(2)

AFOSR-TR- 89-0079

Interface Foundation of North America

20th Symposium on the Interface: Computing Science and Statistics

Theme: Computationally Intensive Methods in Statistics

April 20-23, 1988

Final Report

AD-A205 068

INTERFACE '88

Sponsors

George Mason University
Air Force Office of Scientific Research
Army Research Office
National Bureau of Standards
National Science Foundation
Office of Naval Research

Cooperating Societies

American Mathematical Society
American Statistical Association
Institute of Mathematical Statistics
Washington Statistical Society
International Association for Statistical Computing
Society of Industrial and Applied Mathematics
National Computer Graphics Association
Operations Research Society of America
Virginia Academy of Science Chapter of ASA

89 2 15 143

unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 89-0079

6a. NAME OF PERFORMING ORGANIZATION
Interface Foundation of North America, Inc.

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR/NM

6c. ADDRESS *(City, State, and ZIP Code)*
P.O. Box 7460
Fairfax Station, VA 22039-7460

7b. ADDRESS *(City, State, and ZIP Code)*
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL *(If applicable)*
NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR-88-0154

8c. ADDRESS *(City, State, and ZIP Code)*
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304	A5	

11. TITLE *(Include Security Classification)*
20th Symposium on the Interface: Computing Science and Statistics
Final Report

12. PERSONAL AUTHOR(S)
Edward J. Wegman and Jan P. Guenther

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 4/1/88 TO 8/20/88

14. DATE OF REPORT *(Year, Month, Day)*
1988 August 20

15. PAGE COUNT
195

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Computing Science, Statistics, Computational Statistics Computationally Intensive, Bootstrapping, Parallel Computing, Supercomputing, Neural Networks. (JES)

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The 20th Symposium on the Interface: Computing Science and Statistics was held 20-23 April 1988 in Reston, VA. The theme was computationally intensive methods in statistics. Some 60 invited papers, 128 contributed papers were presented to 425 attendees. There was a special focus on young investigators. Sessions were organized into some 49 technical sessions. This report highlights those sessions and presents abstracts of most of the presentations. Also included is a list of attendees and detailed accounting of expenses.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Maj. Brian Woodruff

22b. TELEPHONE *(Include Area Code)*
202-767-5027

22c. OFFICE SYMBOL
NM

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

DTIC COPY INSPECTED 6

Interface '88
Symposium on the Interface: Computing Science and Statistics

Final Report

prepared by
Edward J. Wegman
Program Chairman of Interface '88
P. O. Box 7460
Fairfax Station, VA 22039-7460

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

Interface '88, the 20th Symposium on the Interface: Computing Science and Statistics, was the first of the Interface Symposia held under the auspices of the Interface Foundation of North America, a non-profit, educational corporation. The Symposium was extremely successful. The attached program and abstracts indicate the quality and scope of the meeting. There were approximately 130 contributed papers up from approximately 60 the prior year. There were some 60 invited papers up somewhat from the previous year. Attendance jumped from about 300 to about 425. We received numerous compliments on the organization and the quality of the program.

Some highlights and innovations we feel pleased to report. For the first time, Interface '88 had a series of special invited papers along with the plenary address. Professor Bradley Efron gave the plenary address. Professors Jerry Friedman, George Box and Tom Banchoff were the three special invited lecturers. These sessions proved to be extremely well attended (to overflow crowds) and sharpened the focus of the meeting. We also introduced for the first time a special invited session for new Ph.D.'s to focus attention on their research. Other sessions which were new to this meeting included sessions on Discrete Mathematics, Symbolic Computation, Supercomputing, Neural Networks and Object Oriented Programming. An emerging area which received attention in the contributed sessions was on Information Systems, Databases and Statistics. This meeting was also the first to have a serious technical focus which was Computationally Intensive Statistical Methods. Keywords

The exhibits were by invitation only. The exhibitors were invited on the basis of their ability to complement the technical program. Additional cooperating societies were involved in Interface '88. New with this meeting were the American Mathematical Society, the National Computer Graphics Association, the Operations Research Society of America, The Washington Statistical Society and the Virginia Academy of Science's Chapter of ASA. This year with the help of the funding agencies, we introduced a young investigator's fund used primarily to fund young Ph.D.'s and graduate student attendance at the Interface. More than $10,000 was set aside for this purpose. This was a highly successful and well received innovation.

Interface '89 is scheduled for Orlando Florida in early April. The University of Central Florida is the host institution with local arrangements being made by Professor Linda Malone. Professor Ken Berk of Illinois State University is the Program Chairman. Interface '90 will be held in East Lansing at Michigan State University. Professor Raoul LePage will be the Program Chairman. Interface '91 will be under the Chairmanship of Dr. John Kettenring of Bell Communications Research. The site will likely be on the West Coast, but final arrangements have yet to be made.

This final report is organized as follows: Immediately following in Appendix A is the Program Information, Program Schedule and Abstracts. Appendix B contains the detailed list of paid attendees. As can be expected, some attendees failed to pay registration fees and hence are not recorded. We believe actual attendance was closer to 445. Appendix C contains the detailed expenditures billed to the Air Force Office of Scientific Research.

Appendix A
Program Information, Program Schedule
Abstracts and Participant Index

Symposium Chairman

Edward J. Wegman
Center for Computational Statistics
George Mason University
Fairfax, VA 22030
(703) 323-2723
EMAIL: EWEGMAN@GMUVAX (bitnet) or
EWEGMAN@GMUVAX.GMU.EDU (arpanet)

Symposium Coordinator and Exhibit Manager

Jan P. Guenther
Center for Computational Statistics
George Mason University
Fairfax, VA 22030
(703) 764-6170

Program Committee

Past Interface Symposia

Southern California, 1968, 1969, 1970, 1971	Chairs: Arnold Goodman, Nancy Mann
Oklahoma State University, 1972 5th Symposium	Chair: Mitchell O. Locks Keynote Speaker: H. O. Hartley
University of California, Berkeley, 1973 6th Symposium	Chair: Michael Tarter Keynote Speaker: John Tukey
Iowa State University, 1974 7th Symposium	Chair: William J. Kennedy Keynote Speaker: Martin Wilk
University of California, Los Angeles, 1975 8th Symposium	Chair: James W. Frane Keynote Speaker: Edwin Kuh
Harvard University, 1976 9th Symposium	Chairs: David Hoaglin and Roy E. Welsch Keynote Speaker: John R. Rice
National Bureau of Standards, 1977 10th Symposium	Chair: David Hogben Keynote Speaker: Anthony Ralston
North Carolina State University, 1978 11th Symposium	Chairs: Ron Gallant and Thomas Gerig Keynote Speaker: Nancy Mann
University of Waterloo, 1979 12th Symposium	Chair: Jane F. Gentleman Keynote Speaker: D. R. Cox
Carnegie-Mellon University, 1981 13th Symposium	Chair: William F. Eddy Keynote Speaker: Brad Efron
Rensselaer Polytechnic Institute, 1982 14th Symposium	Chairs: John W. Wilkinson, Karl W. Heiner and Richard Sacher Keynote Speaker: John Tukey
IMSL, Inc (held in Houston), 1983 15th Symposium	Chair: James Gentle Keynote Speaker: Richard Hamming
University of Georgia (held in Atlanta), 1984 16th Symposium	Chair: Lynne Billard Keynote Speaker: George Marsalgia
University of Kentucky, 1985 17th Symposium	Chair: David Allen Keynote Speaker: John C. Nash
Colorado State University, 1986 18th Symposium	Chair: Thomas Boardman Keynote Speaker: John Tukey

Past Interface Symposia (Continued)

Temple University (held in Philadelphia), 1987 19th Symposium	Chair: Richard Heiberger Keynote Speaker: Gene Golub
George Mason University, 1988 20th Symposium	Chair: Edward J. Wegman Keynote Speaker: Brad Efron

Future Interface Symposia

University of South Florida, 1989 21st Symposium	Chairs: Ken Berk and Linda Malone
Michigan State University, 1990 22nd Symposium	Chair: Raoul LePage

General Information

The 20th Symposium represents a milestone in the development of the interface between computing science and statistics. In August, 1987 the Interface Foundation of North America was incorporated as a non-profit, educational corporation whose main charter is to provide the legal entity underpinning the Symposium series. The Foundation represents a maturation of the Symposium series and ensures its continuation as an independent meeting focused on the interface. The 20th Symposium is the first held under the auspices of the Foundation. It is also the first with a focused theme.

Theme: – Computationally Intensive Statistical Methods

Keynote Address: – "Computationally intensive statistical inference"
Bradley Efron, Department of Statistics, Stanford University

Invited Papers: – There are 60 invited papers including several with invited discussion organized into 23 sessions. In addition to the plenary session with the keynote address by Brad Efron, there are three special invited lectures featuring Jerome Friedman, George E. P. Box and Thomas Banchoff.

Contributed Papers: – There are 128 contributed papers scheduled in 26 sessions.

Proceedings: – The Proceedings of the 20th Interface Symposium will be published by the American Statistical Association and will be available late autumn of 1988.

Opening Reception: – All registrants are invited to attend the Opening Reception on Wednesday evening from 8:00 p.m. until 10:00 p.m. The Reception will include a light food service and two tickets for drinks will be provided registrants. A cash bar will be available thereafter. The Reception will be held in the hotel ballroom.

Banquet: – The Banquet will be served buffet style on Friday evening beginning at 7:00 p.m. The planned menu includes roast turkey, baked ham, seafood in leek and wine sauce, roast beef, and chicken in almond sauce. The banquet is a separate cost item. It will be held in the hotel ballroom following a cash bar beginning at 6:00 p.m. Following the banquet, the Mill Run Dulcimer Band, a Washington-area based bluegrass group will perform. As many may known, the Washington, D. C. area is noted as a headquarters area for bluegrass and old-time country music.

Other Food Service: – Coffee and Danish will be served during the Thursday, Friday and Saturday morning breaks and soft drinks and cookies during the afternoon breaks on Thursday and Friday. These food services will be available in the exhibit area. Luncheons and other meals will be at the option of the registrants and may be obtained in the hotel or in nearby restaurants. A cash bar will also be available on Thursday evening from 6:00 p.m. until 9:00 p.m.

Shuttle Service: – A free shuttle service is provided by the hotel to and from the Dulles International Airport on the half hour. In addition, the hotel will be running a shuttle service to and from the Vienna Metro (subway) station. The schedule of service will be posted. The Metro systems provides convenient and economical access to the downtown Washington metropolitan area.

Exhibits: – The exhibit area is located in rooms 9 and 10 of the hotel. Exhibits will be available to registrants immediately following the Plenary session on Thursday morning through the close of the Symposium on Saturday.

Exhibitors

Ametek Computer Corporation
606 East Huntington Drive
Monrovia, CA 91016
(714) 599-4662

Automatic Forecasting Systems, Inc.
P. O. Box 563
Hatboro, PA 19040
(215) 675-0652

BBN Software
10 Fawcett Street
Cambridge, MA 02238
(617) 873-8116

BMDP Statistical Software, Inc.
1440 Sepulveda Boulevard, Suite 316
Los Angeles, CA 90025
(213) 479-7799

Intel Scientific Computers
15201 NW Greenbrier Parkway
Beaverton, OR 97006
(503) 629-7631

Marcel-Dekker, Inc..
270 Madison Avenue
New York, NY 10016
(212) 696-9000

IMSL, Inc.
2500 ParkWest Tower One
2500 CityWest Boulevard
Houston, TX 77042-3020
(713) 782-6060

North Holland/Elsevier Publishers
P. O. Box 1991
1000 BZ Amsterdam
The Netherlands

Numerical Algorithms Group
1101 31st Street, Suite 100
Downers Grove, IL 60515
(312) 971-2337

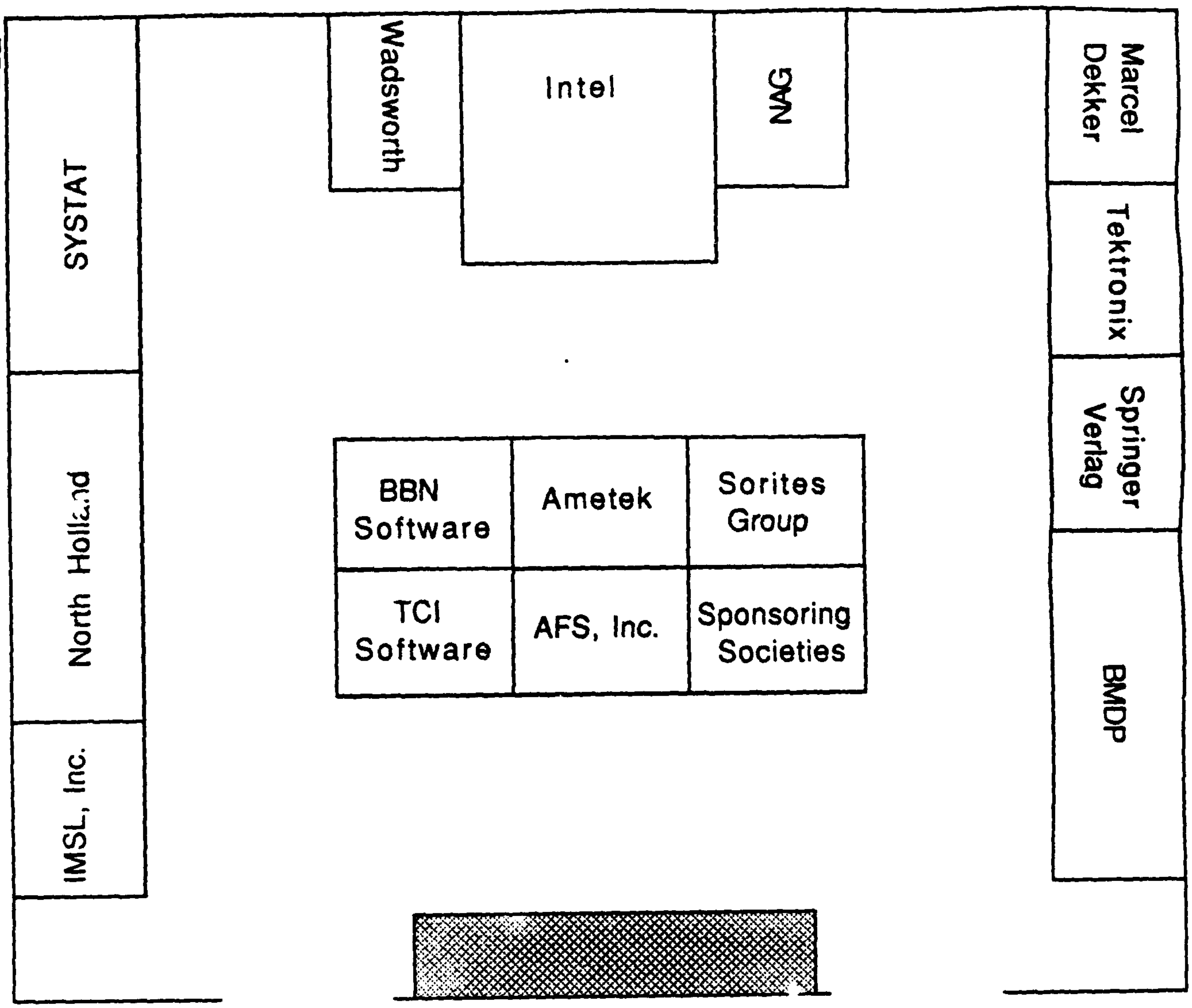
Springer-Verlag, Inc.
175 Fifth Avenue
New York, NY 10010
(212) 460-1600

SYSTAT, Inc.
1800 Sherman Avenue
Evanston, IL 60201
(312) 864-5670

TCI Software
1190 Foster Road
Las Cruces, NM 88001
(505) 522-4600

Tektronix, Inc.
M.S. 48-300, Industrial Park
Beaverton, OR 97077
(503) 627-7111

Wadsworth & Brooks/Cole
Advanced Books and Software
10 Davis Drive
Belmont, CA 94002
(415) 595-2350



SYSTAT
North Holland
IMSL, Inc.
Wadsworth
Intel
NAG
Marcel Dekker
Tektronix
Springer Verlag
BMDP
BBN Software
Ametek
Sorites Group
TCI Software
AFS, Inc.
Sponsoring Societies

Short Course

Forecasting on the IBM-PC - A Survey, Wednesday, April 20, 9:00 a.m. to 4:30 p.m., David P. Reilly, Automatic Forecasting Systems, Inc., P. O. Box 563, Hatboro, PA 19040, (215) 675-0652

Cooperating Societies

American Mathematical Society
P. O. Box 6248
Providence, RI 02940

American Statistical Association
1429 Duke Street
Alexandria, VA 22314

International Association for Statistical Computing
NTDH
P. O. Box 145
N-7701 Steinkjer
Norway

Institute of Mathematical Statistics
3401 Investment Boulevard, Suite 7
Hayward, CA 94545

National Computer Graphics Association
2722 Merilee, Suite 200
Fairfax, VA 22031

Operations Research Society of America
Mount Royal and Guilford Avenues
Baltimore, MD 21202

Society for Industrial and Applied Mathematics
1400 Architects Building
117 South 17th Street
Philadelphia, PA 19103

Virginia Academy of Sciences Chapter of the ASA
c/o Golde I. Holtzman
Department of Statistics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

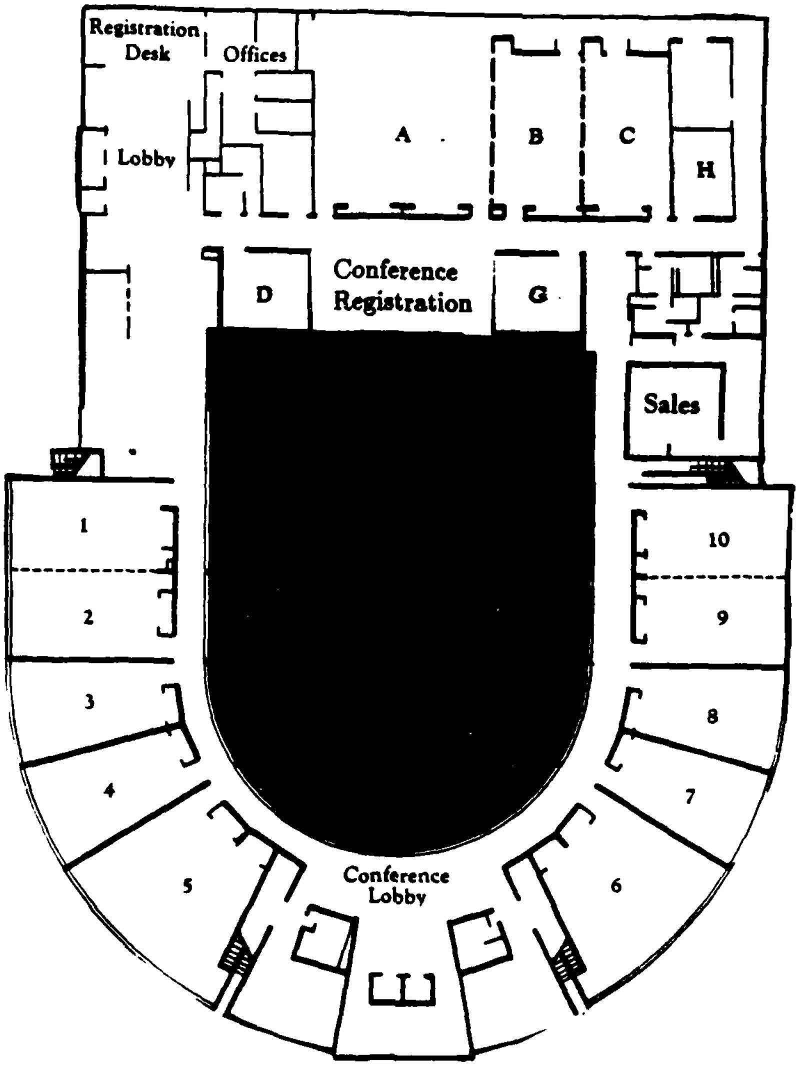
Washington Statistical Society
P. O. Box 70843
Washington, DC 20024-0843

Program Schedule

Date and Time	Session Title	Room
Thursday, April 21		
8:45 a.m. - 9:45 a.m.	Keynote Address: Computationally Intensive Statistical Inference	Ballroom
10:00 a.m. - 12:00 noon	Computational Aspects of Time Series Analysis	Room 6
	Inference and Artificial Intelligence	Room 5
	Computational Discrete Mathematics	Room 3
	Contributed: Software Tools	Room 2
	Contributed: Image Processing I	Room D
	Contributed: Bootstapping and Related Computational Methods	Room 1
1:30 p.m. - 3:30 p.m.	Special Invited Lecture I	Room 6
	Image Processing and Spatial Processes	Room 5
	Parallel Computing Architectures	Room 3
	Contributed: Statistical Methods I	Room 2
	Contributed: Hardware and Software Reliability	Room D
	Contributed: Applications I	Room 1
3:45 p.m. - 5:45 p.m.	Special Invited Session for Recent Ph.D.'s	Room 6
	Simulation	Room 5
	Symbolic Computation and Statistics	Room 3
	Contributed: Statistical Graphics	Room 2
	Contributed: Models of Imprecision in Expert Systems	Room D
	Contributed: Time Series Methods	Room 1
Friday, April 22		
8:00 a.m. - 10:00 p.m.	Computer-Communication Networks	Room 6
	Supercomputing, Design of Experiments and Bayesian Analysis, Part 1	Room 5
	Numerical Methods in Statistics	Room 3
	Contributed: Probability and Stochastic Processes	Room 2
	Contributed: Statistical Methods II	Room D
	Contributed: Nonparametric and Robust Techniques	Room 1
10:15 a.m. - 12:15 p.m.	Special Invited Lecture II	Room 6
	Supercomputing, Design of Experiments and Bayesian Analysis, Part 2	Room 5
	Neural Networks	Room 3
	Contributed: Applications II	Room 2
	Contributed: Image Processing II	Room D
	Contributed: Simulation I	Room 1

2:00 p.m. - 4:00 p.m.	Tales of the Unexpected: Successful Interdisciplinary Research	Room 6
	Density Estimation and Smoothing	Room 5
	Object Oriented Programming	Room 3
	Contributed: Numerical Methods	Room 2
	Contributed: Bayesian Methods	Room D
	Contributed: Expert Systems in Statistics	Room 1
Saturday, April 23		
8:30 a.m. - 10:30 a.m.	Computational Aspects of Simulated Annealing	Room 6
	Dynamical High Interaction Graphics	Room 5
	Contributed: Statistical Methods III	Room 3
	Contributed: Simulation II	Room 2
	Contributed: Biostatistics Applications	Room D
	Contributed: Discrete Mathematical Methods	Room 1
10:45 a.m. - 12:45 p.m.	Special Invited Lecture III	Room 6
	Entropy Methods	Room 5
	Contributed: Information Systems, Databases and Statistics	Room 3
	Contributed: Parallel Computing	Room 2
	Contributed: Density and Function Estimation	Room D
	Contributed: Statistical Methods IV	Room 1

Conference Center Layout



Thursday, April 21, 1988

8:45 - 9:45	**Plenary Session**					
	Invited			Contributed		
10:00-12:00	Computational Aspects of Time Series Analysis	Inference and Artificial Intelligence	Computational Discrete Mathematics	Software Tools	Image Processing I	Bootstrapping and Related Computational Methods
1:30-3:30	Special Invited Lecture I	Image Processing and Spatial Processes	Parallel Computing Architecture	Statistical Methods I	Hardware and Software Reliability	Applications I
3:45-5:45	Special Invited Session for Recent Ph.D's	Simulation	Symbolic Computation and Statistics	Statistical Graphics	Models of Imprecision in Expert Systems	Time Series Methods

Friday, April 22, 1988

	Invited			Contributed		
8:00-10:00	Computer Communication Networks	Supercomputing, Design of Experiments, and Bayesian Analysis, Pt.1	Numerical Methods In Statistics	Probability and Stochastic Processes	Statistical Methods II	Nonparametric and Robust Techniques
10:15-12:15	Special Invited Lecture II	Supercomputing, Design of Experiments, and Bayesian Analysis, Pt.2	Neural Networks	Applications II	Image Processing II	Simulation I
2:00-4:00	Tales of the Unexpected: Successful Interdisciplinary Research	Density Estimation and Smoothing	Object Oriented Programming	Numerical Methods	Bayesian Methods	Expert Systems in Statistics

Saturday, April 23, 1988

	Invited		Contributed			
8:30-10:30	Computational Aspects of Simulated Annealing	Dynamical High Interaction Graphics	Statistical Methods III	Simulation II	Biostatistics Application	Discrete Mathematical Methods
10:45-12:45	Special Invited Lecture III	Entropy Methods	Information Systems, Databases and Statistics	Parallel Computing	Density and Function Estimation	Statistical Methods IV

Technical and Social Program

WEDNESDAY, APRIL 20, 1988

9:00 a.m. - 4:30 p.m. — Room 6
Short Course - Forecasting on the IBM-PC, David Reilly, Automatic Forecasting Systems, Inc.

4:00 p.m. — Lobby
Registration for Symposium

5:00 p.m. — Room G
Interface Board of Directors Meeting (by invitation only)

8:00 p.m. - 10:00 p.m. — Ballroom
Free Opening Reception

THURSDAY, APRIL 21, 1988

7:30 a.m. — Lobby
Registration

8:30 a.m. - 8:45 a.m. — Ballroom
Welcoming Remarks

8:45 a.m. - 9:45 a.m. — Ballroom
Plenary Session, Chaired by: Edward J. Wegman, George Mason University

"Computationally intensive statistical inference," Bradley Efron, Stanford University

10:00 a.m. - 12:00 noon — Room 6
Computational Aspects of Time Series Analysis, Chaired by: Emanuel Parzen, Texas A & M University

"Recent progress in algorithms and architectures for time series analysis," George Cybenko, Tufts University

"Numerical approach to non-gaussian smoothing and its application," Genshiro Kitagawa, The Institute of Statistical Mathematics

Discussants - Will Gersch, University of Hawaii and H. Joseph Newton, Texas A & M University

THURSDAY, APRIL 21, 1988

10:00 a.m. - 12:00 noon **Room 5**

Inference and Artificial Intelligence, Chaired by: N. Singpurwalla, George Washington University

"Spectral Analysis on a LISP machine," Don Percival, University of Washington

"DeFinetti's approach to group decision making," Richard Barlow, University of California, Berkeley

"Meta-analysis," Ingram Olkin, Stanford University

10:00 a.m. - 12:00 noon **Room 3**

Computational Discrete Mathematics, Chaired by: Rich Ringeisen, Clemson University

"Discrete structures and reliability computations," James P. Jarvis, Clemson University and Douglas R. Shier, College of William and Mary

"Random graphs," Edward R. Scheinerman, The Johns Hopkins University

"Structure and finiteness conditions on graphs," Neil Robertson, Ohio State University

10:00 a.m. - 12:00 noon **Room 2**

Contributed Papers: Software Tools, Chaired by: Leonard Hearne, George Mason University

"An introduction to CART^{tm}: classification and regression trees," Gerard T. LaVarnway, Norwich University

"Noise appreciation: analyzing residuals using RS/Explore," David A. Burn and Fanny O'Brien, BBN Software Products Corporation

"COSTAR: an environment for computer-guided data analysis," David A. Whitney and Ilya Schiller, TASC

"A closer look at symbolic computation," William M. Makuch, General Electric Corporation and John W. Wilkinson, Rensselaer Polytechnic Institute

10:00 a.m. - 12:00 noon **Room D**

Contributed Papers: Image Processing I, Chaired by: A. K. Sood, George Mason University

"Image analysis of a turbulent object using fractal parameters," Amar Ait-Kheddache, North Carolina State University

"Identification of closed figures," Jeff Banfield, Montana State University and Adrian Raftery, University of Washington

"Compression of image data using arithmetic coding," Ahmed H. Desoky and Thomas Klein, University of Louisville

THURSDAY, APRIL 21, 1988

"Image analysis of the microvascular system in the rat cremaster muscle," C. O'Connor, P. D. Harris, A. Desoky and G. Ighodaro, University of Louisville

"Automatic detection of the optic nerve in color images of the retina," Norman Katz, Subhasis Chaudhuri, and Michael Goldbaum, University of California, San Diego and Mark Nelson, Radford Company

10:00 a.m. - 12:00 noon **Room 1**

Contributed Papers: Bootstrapping and Related Computational Methods, Chaired by: Richard Bolstein, George Mason University

"A Monte Carlo study of cross-validation and the C_p criterion for model selection in multiple linear regression," Robert M. Boudreau, Virginia Commonwealth University

"Bootstrapping regression strategies," David Brownstone, University of California, Irvine

"Bootstrapping the missed regression model with reference to the capital and energy complementarity debate," Baldev Raj, Wilfred Laurier University

"Efficient data sensitivity computation for maximum likelihood estimation," Daniel Chin and James C. Spall, The Johns Hopkins University

"Bootstrap procedures in random effect models for comparing response rates in multi-center clinical trials," Michael F. Miller, Hoechst-Roussel Pharmaceuticals, Inc.

1:30 p.m. - 2:45 p.m. **Room 6**

Special Invited Lecture I, Chaired by: Jim Filliben, National Bureau of Standards

"Fitting functions to scattered noisy data in high dimensions," Jerome Friedman, Stanford University

1:30 p.m. - 3:30 p.m. **Room 5**

Image Processing and Spatial Processes, Chaired by: Don McClure, Brown University

Introduction, Don McClure, Brown University

"A multilevel-multiresolution technique for image analysis and robot vision via renormalization group ideas," Basilis Gidas, Brown University

"A mathematical approach to expert system construction," Alan Lippman, Brown University

THURSDAY, APRIL 21, 1988

1:30 p.m. - 3:30 p.m. Room 3

Parallel Computing Architectures, Chaired by: Chris Brown, University of Rochester

"Experiences with the BBN Butterfly™ parallel processor," John Mellor-Crummy, University of Rochester

"Statistical computing on a hypercube," George Ostrouchov, Oak Ridge National Lab

"Asychronous iteration," William F. Eddy and Mark Schervish, Carnegie-Mellon University

1:30 p.m. - 3:30 p.m. Room 2

Contributed Papers: Statistical Methods I, Chaired by: Walter Liggett, National Bureau of Standards

"An example of the use of a Bayesian interpretation of multiple discriminant analysis results," James R. Nolan, Siena College

"Real-time classification and discrimination among components of a mixture distribution," Douglas A. Samuelson, International Telesystems Corporation

"Comparison of three 'local model' classification methods," Daniel Normolle, University of Michigan

"Application of posterior approximation techniques for the ordered Dirichlet distribution," Thomas A. Mazzuchi and Refik Soyer, George Washington University

"Unbiased estimates of multivariate general moment functions of the population and application to sampling without replacement for a finite population," Nabih N. Mikhail, Liberty University

1:30 p.m. - 3:30 p.m. Room D

Contributed Papers: Hardware and Software Reliability, Chaired by: Asit Basu, University of Missouri

"Linear prediction of failure times of a repairable system," M. Ahsanullah, Rider College

"The simulation of life tests with random censoring," Joseph C. Hudson, GMI Engineering and Management Institute

"The use of general modified exponential curves in software reliability modeling," Taghi M. Khoshgoftaar, Florida Atlantic University

"A model for information censoring," William A. Link, Patuxent Wildlife Research Center

"Increasing reliability of multiversion fault-tolerant software design by modulation," Junryo Miyashita, California State University, San Bernardino

THURSDAY, APRIL 21, 1988

1:30 p.m. - 3:30 p.m. **Room 1**

Contributed Papers: Applications I, Chaired by: Susannah Schiller, National Bureau of Standards

"Classifying linear mixtures with an application to high resolution gas chromatography," William S. Rayens, University of Kentucky

"Bias of animal trend estimates," Paul H. Geissler and William A. Link, Patuxent Wildlife Research Center

"A non-random walk through futures prices of the British pound," William S. Mallios, California State University, Fresno

"A stochastic extension of Petri net graph theory," L. M. Anneberg, Wayne State University

"Neural Petri nets," N. H. Chamas, Wayne State University

3:45 p.m. - 5:45 p.m. **Room 6**

Special Invited Session for Recent Ph.D.'s, Chaired by: John J. Miller, George Mason University

"Additive principal components: a method for estimating equations with small variance from multivariate data," Deborah Donnell, Bellcore

"Gamma processes, paired comparisons and ranking," Hal Stern, Harvard University

"Smoothing data with correlated errors," Naomi Altman, Cornell University

"The data viewer: program for graphical data analysis," Catherine Hurley, University of Waterloo

3:45 p.m. - 5:45 p.m. **Room 5**

Simulation, Chaired by: Donald T. Gantz, George Mason University

"Random variables for supercomputers," George Marsaglia, Florida State University

"Computational statistics in experimental design for studies of variability," John Ramberg, University of Arizona

"Linear combinations of estimators of the variance of the sample mean," Bruce W. Schmeiser, Purdue University

THURSDAY, APRIL 21, 1988

3:45 p.m. - 5:45 p.m. Room 3

Symbolic Computation and Statistics, Chaired by: William S. Rayens, University of Kentucky

"Some applications of symbol manipulation in statistical analysis," Kathryn M. Chaloner, University of Minnesota

"Symbolic computation in statistical decision theory," Marietta Tretter, Texas A & M University

"Partial differentiation by computer with applications to statistics," John W. Sawyer, Jr., Texas Tech University

3:45 p.m. - 5:45 p.m. Room 2

Contributed Papers: Statistical Graphics, Chaired by: Robert Launer, Army Research Office

"Visual multidimensional geometry with applications," Alfred Inselberg, IBM Scientific Center, Los Angeles and Bernard Dimsdale, University of California

"Some graphical representations of multivariate data," Masood Bolorforoush and Edward J. Wegman, George Mason University

"Graphical representations of main effects and interaction effects in a polynomial regression on several predictors," William DuMouchel, BBN Software Products Corporation

"Chernoff faces: a PC implementation," Mohammad Dadashzadeh, University of Detroit

3:45 p.m. - 5:45 p.m. Room D

Contributed Papers: Models of Imprecision in Expert Systems, Chaired by: Mark Youngren, George Washington University

"Fusion and propagation of graphical belief models," Russell Almond, Harvard University

"Belief function computations for paired comparisons," David Tritchler and Gina Lockwood, Ontario Cancer Institute

"Variants of Tierney-Kadane," Guenter Weiss and H. A. Howlader, University of Winnepeg

"Dynamically updating relevance judgements in probabilistic information systems via users' feedback," Peter Lenk and Barry D. Floyd, New York University

"Computational requirements for inference methods in expert systems: a comparative study," Ambrose Goicoechea, George Mason University

THURSDAY, APRIL 21, 1988

3:45 p.m. - 5:45 p.m. Room 1

Contributed Papers: Time Series Methods, Chaired by: Neil Gerr, Office of Naval Research

"Inference techniques for a class of exponential time series," V. Chandrasekar and Peter Brockwell, Colorado State University

"Some recursive methods in time series analysis," Q. P. Duong, Bell Canada

"Time series in a microcomputer environment," John Henstridge, Numerical Algorithms Group

"Smoothing irregular time series," Keith W. Hipel, University of Waterloo, A. I. McLeod, The University of Western Ontario and Byron Bodo, Ministry of the Environment

"Computation of the theoretical autocovariance function of multivariate ARMA processes," Stefan Mittnik, SUNY at Stony Brook

6:00 p.m. - 8:00 p.m. Room G

Executive Session of Statistical Computing Section of ASA (by invitation only)

6:00 p.m. - 9:00 p.m. Ballroom

Cash Bar

FRIDAY, APRIL 22, 1988

8:00 a.m. - 10:00 a.m. Room 6

Computer-Communication Networks, Chaired by: Martin Fischer, Defense Communication Engineering Center

"Introduction to packet switching networks," Jeffrey Mayersohn, BBN Communication Corporation

"Electronic mail - a valuable augmentation tool for scientists," Elizabeth Feinler, SRI International

"Networks to support science," Stephen Wolff, National Science Foundation

FRIDAY, APRIL 22, 1988

8:00 a.m. - 10:00 a.m. Room 5

Supercomputing, Design of Experiments and Bayesian Analysis, Part I, Chaired by: Jerry Sacks, University of Illinois

"Acceleration methods for Monte Carlo integration by Bayesian inference," John Geweke, Duke University

"Software for Bayesian analysis: current status and additional needs," Prem K. Goel, Ohio State University

"Some numerical and graphical stategies for implementing Bayesian methods," Adrian Smith, University of Nottingham

8:00 a.m. - 10:00 a.m. Room 3

Numerical Methods for Statistics, Chaired by: Stephen Nash, George Mason University

"Interior point methods for linear programming," Paul Boggs, National Bureau of Standards

"Block iterative methods for parallel optimization," Stephen Nash and Ariela Sofer, George Mason University

"New methods for B-differentiable functions: theory and applications," Jong-Shi Pang, The Johns Hopkins University

8:00 a.m. - 10:00 a.m. Room 2

Contributed Papers: Probability and Stochastic Processes, Chaired by: Yash Mittal, National Science Foundation

"Moving window detection for 0-1 Markov trials," Joseph Glaz, University of Connecticut, Philip C. Hormel, CIBA-GEIGY Corporation and Bruce McK. Johnson, University of Connecticut

"Maximum queue size and hashing with lazy deletion," Claire M. Mathieu, Laboratoire d'Informatique de l'Ecole Normale Superieure and Jeffrey S. Vitter, Brown University

"On the probability integrals of the multivariate normal," Dror Rom and Sanat Sarkar, Temple University

"Computational aspects of harmonic signal detection," Keh-Shin Lii and Tai-Houn Tsou, University of California, Riverside

"Maximum likelihood estimation of discrete control processes: theory and application," John Rust, University of Wisconsin

FRIDAY, APRIL 22, 1988

8:00 a.m. - 10:00 a.m. **Room D**

Contributed Papers: Statistical Methods II, Chaired by: Cliff Sutton, George Mason University

"Computing extended maximum likelihood estimates in generalized linear models," Douglas B. Clarkson, IMSL, Inc. and Robert I. Jennrich, University of California, Los Angeles

"Assessment of prediction procedures in multiple regression analysis," Victor Kipnis, University of Southern Florida

"Estimation of the variance matrix for maximum likelihood parameters by quasi-Newton methods," Linda Pickle, National Cancer Institute and Garth P. McCormick, George Washington University

"Variable selection in multivariate multiresponse permutation procedures," Eric P. Smith, Virginia Tech

"The effect of small covariate-criterion correlations on analysis of covariance," Michael J. Rovine, A. von Eye and P. Wood, Pennsylvania State University

8:00 a.m. - 10:00 a.m. **Room 1**

Contributed Papers: Nonparametric and Robust Techniques, Chaired by: Paul Speckman, University of Missouri

"Robustness of weighted estimators of location: a small sample survey," Greg Campbell and Richard I. Shrager, NIH

"A comparison of Spearman's footrule and rank correlation coefficient with exact tables and approximations," LeRoy A. Franklin, Indiana State University

"Approximations of the Wilcoxon test in small samples with lots of ties," Arthur R. Silverberg, Food and Drug Administration

"Simulated power comparisons of MRPP rank tests and some standard score tests," Derrick S. Tracy and Khushnood A. Khan, University of Windsor

10:15 a.m. - 12:15 p.m. **Room 6**

Special Invited Lecture II, Chaired by: Mervin Muller, Ohio State University

"Some modern quality improvement techniques and their computing implications," George E. P. Box, University of Wisconsin

Special invited discussion, Gerald J. Hahn, GE CRD and Gregory B. Hudak, Scientific Computing Associates

FRIDAY, APRIL 21, 1988

10:15 a.m. - 12:15 p.m. **Room 5**

Supercomputing, Design of Experiments and Bayesian Analysis, Part II, Chaired by: Prem K. Goel, Ohio State University

"Supercomputer-aided design," Jerry Sacks, University of Illinois

"A Bayesian approach to the design and analysis of computer experiments," Toby Mitchell, Oak Ridge National Lab

10:15 a.m. - 12:15 p.m. **Room 3**

Neural Networks, Chaired by: Muhammed Habib, University of North Carolina

"Statistical learning networks: a unifying view," Andrew R. Barron, University of Illinois and Roger L. Barron, Barron Associates, Inc.

"Stochastic models of neuronal behavior," Gopinath Kallianpur, University of North Carolina

"Inference for stochastic models for neural networks," Muhammed Habib, University of North Carolina and A. Thavaneswaran, Temple University

10:15 a.m. - 12:15 p.m. **Room 2**

Contributed Papers: Applications II, Chaired by: **Brian Woodruff, Air Force Office of Scientific Research**

"Space Balls! or estimating diameter distributions of polystyrene microspheres," Susannah Schiller and Charles Hagwood, National Bureau of Standards

"Comparing sample reuse methods at FHA - an empirical approach," Thomas N. Herzog, U. S. Department of Housing and Urban Development

"Maximum entropy and its application to linguistic diversity," R. K. Jain, Memorial University of Newfoundland

"Encoding and processing of Chinese language - a statistical structural approach," Chaiho C. Wang, George Washington University

"The elimination of quantization bias using dither," Martin J. Garbo and Douglas M. Dreher, Hughes Aircraft Company

10:15 a.m. - 12:15 p.m. **Room D**

Contributed Papers: Image Processing II, Chaired by: **Refik Soyer, George Washington University**

"Maximum entropy and the nearly black image," Iain Johnstone, Stanford University and David Donoho, University of California, Berkeley

"A probabilistic approach to range image description," Arun Sood, George Mason University and E. Al-Hujazi, Wayne State University

FRIDAY, APRIL 22, 1988

"An empirical Bayes decision rule of two-class pattern recognition for one-dimensional parametric distributions," Tze Fen Li, Rutgers University

"Statistical modeling of a priori information for image processing problems," Z. Liang, Duke University Medical Center

"Advanced statistical computations improve image processing applications, Bobby Saffari, Generex Corporation

10:15 a.m. - 12:15 p.m. Room 1

Contributed Papers: Simulation I, Chaired by: Bill DuMouchel, BBN

"On comparative accuracy of multivariate nonnormal random number generators," Lynne K. Edwards, University of Minnesota

"Bayesian analysis using Monte Carlo integration: an effective methodology for handling some difficult problems in statistical analysis," Leland Stewart, Lockheed Research Laboratory

"A squeeze method for generating exponential power variates," Dean M. Young, Baylor University

"Mixture experiments and fractional factorials used to tailor large-scale computer simulation," T.K. Gardenier, TKG Consultants, Ltd.

"Simulating stationary Gaussian ARMA time series," Terry J. Woodfield, SAS Institute, Inc.

2:00 p.m. - 4:00 p.m. Room 6

Tales of the Unexpected: Successful Interdisciplinary Research, Chaired by: Sallie McNulty, Kansas State University

"Some statistical problems in meteorology," Grace Wahba, University of Wisconsin

"Modeling parallelism, an interdisciplinary approach," Elizabeth Unger, Kansas State University

"Mice, rain forests and finches: experiences collaborating with biologists," Douglas Nychka, North Carolina State University

Discussion: Jerome Sacks, University of Illinois

FRIDAY, APRIL 22, 1988

2:00 p.m. - 4:00 p.m. Room 5

Density Estimation and Smoothing, Chaired by: David Scott, Rice University

"XploRe: computing environment for exploratory regression and density estimation methods," Wolfgang Härdle, University of Bonn

"Curve estimation with applications to mapping and risk decomposition," Michael Tarter, University of California, Berkeley

"Interactive multivariate density estimation in the S package," David Scott, Rice University

2:00 p.m. - 4:00 p.m. Room 3

Object Oriented Programming, Chaired by: Werner Stuetzle, University of Washington

"Object oriented programming: a tutorial," Wayne Oldford, University of Waterloo

"An object oriented toolkit for plotting and interface construction," Robert Young, Schlumberger, Palo Alto Research Center

"An outline of Arizona," John MacDonald, University of Washington

2:00 p.m. - 4:00 p.m. Room 2

Contributed Papers: Numerical Methods, Chaired by: Ariela Sofer, George Mason University

"A theorgy of quadrature in applied probability: a fast algorithmic approach," Allen Don, Long Island University

"Higher order functions in numerical programming," David Gladstein, ICAD

"A numerical comparison of EM and quasi-Newton type algorithms for finding MLE's for a mixture of normal distributions," Richard J. Hathaway, John W. Davenport and Margaret Anne Pierce, Georgia Southern College

"Numerical algorithms for exact calculations of early stopping probabilities in one-sample clinical trials with censored exponential responses," Brenda MacGibbon, Concordia University, Susan Groshen, University of Southern California and Jean-Guy Levreault, University of Montreal

"An application of quasi-Newton methods in parametric empirical Bayes calculations," David Scott, Concordia University

FRIDAY, APRIL 22, 1988

2:00 p.m. - 4:00 p.m. Room D

Contributed Papers: Bayesian Methods, Chaired by: William F. Eddy, Carnegie-Mellon University

"Approaches for empirical Bayes confidence intervals with application to exponential scale parameters," Alan E. Gelfand and Bradley P. Carlin, University of Connecticut

"A data analysis and Bayesian framework for errors-in-variables," John H. Herbert, Department of Energy

"Bayesian diagnostics for almost any model," Robert E. Weiss, University of Minnesota

"An iterative Bayes method for classifying multivariate observations," Duane E. Wolting, Acrojet Tech Systems Company

"A Bayesian model of information conbination from noisy sensors," G. Anandalingam, University of Pennsylvania

2:00 p.m. - 4:00 p.m. Room 1

Contributed Papers: Expert Systems in Statistics: Chaired by Khalid Abouri, George Washington University

"Inside a statistical expert system: implementation of the ESTES expert system," Paula Hietala, University of Tampere, Finland

"Knowledge-based project management: work effort estimation," Vijay Kanabar, University of Winnipeg

"Combining knowledge acquisition and classical statistical techniques in the development of a veterinary medical expert system," Mary McLeish, University of Guelph

"The effect of measurement error in a machine learning system," David L. Rumpf and Mieczyslaw M. Kokar, Northeastern University

"An expert system for prescribing statistical tests of non-parametric and simple parametric designs," Gary Tubb, University of South Florida

6:00 p.m. - 7:00 p.m. Ballroom

Cash Bar

7:00 p.m. - 9:30 p.m. Ballroom

Banquet, Live Entertainment (fee event)

SATURDAY, APRIL 23, 1988

8:30 a..m. - 10:30 a.m. **Room 6**

Computational Aspects of Simulated Annealing, Chaired by: Mark E. Johnson, Los Alamos National Lab

"Computational experience with simulated annealing," Daniel G. Brooks and William A. Verdini, Arizona State University

"Simulated annealing in optimal design construction," Ruth K. Meyer, St. Cloud State University and Christopher J. Nachtsheim, University of Minnesota

"A simulated annealing approach to mapping DNA," Larry Goldstein and Michael J. Waterman, University of Southern California

8:30 a.m. - 10:30 a.m. **Room 5**

Dynamical High Interaction Graphics, Chaired by: Paul Tukey, Bellcore

"Determining properties of minimal spanning trees by local sampling," Allen McIntosh, Bellcore and William Eddy, Carnegie-Mellon University

"Data animation," Rick Becker, AT&T Bell Labs and Paul Tukey, Bellcore

"Dimensionality constraints on projection and section views of higher dimensional loci," George Furnas, Bellcore

8:30a.m. - 10:30 a.m **Room 3**

Contributed Papers: Statistical Methods III, Chaired by: Thomas Mazzuchi, George Washington University

"Simultaneous confidence intervals in the general linear model," Jason C. Hsu, Ohio State University

"Empirical likelihood ratio confidence regions," Art Owen, Stanford University

"An approximate confidence interval for the optimal number of mammography x-ray units in the Dallas-Fort Worth metropolitan area," Roger W. Peck, University of Rhode Island

"Optimizing linear functions of random variables having a joint multinomial or multivariate normal distribution," Josephina P. de los Reyes, University of Akron

"On covariances of marginally adjusted data," James S. Weber, Roosevelt University

8:30 a.m. - 10:30 a.m. **Room 2**

Contributed Papers: Simulation II, Chaired by : Robert Jernigan, American University

"SIMDAT and SIMEST: differences and convergences," James R. Thompson, Rice University

"Simulation and stochastic modeling for the spatial allocation of multi-categorical resources," Richard S. Segall, University of Lowell

SATURDAY, APRIL 23, 1988

"Robustness study of some random variate generators," Lih-Yuan Deng, Memphis State University

"Testing multiprocessing random number generators," Mark J. Durst, Lawrence Livermore National Laboratory

"An approach for generations of two variable sets with a specified correlation and first and second sample moments," Mark Eakin and Henry D. Crockett, University of Texas at Arlington

8:30 a.m. - 10:30 a.m. **Room D**

Contributed Papers: Biostatistics Applications, Chaired by: Nancy Flournoy, National Science Foundation

"An algorithm to identify changes in hormone patterns," Morton B. Brown, Fred J. Karsch and Benoit Malpaux, University of Michigan

"Applying microcomputer techniques to multiple cause of death data: from magnetic tape to artificial intelligence," Giles Crane, New Jersey State Department of Health

"Spline estimation of death density using census and vital statistics data," John J. Hsieh, University of Toronto

"Optimum experimental design for sequential clinical trials," Richard Simon, National Cancer Institute

"Bayes estimation of cerebral metabolic rate of glucose in stroke patients," P. David Wilson, University of South Florida, S. C. Huang and R. A. Hawkins, UCLA School of Medicine

8:30 a.m. - 10:30 a.m. **Room 1**

Contributed Papers: Discrete Mathematical Methods, Chaired by: Donald Gants, George Mason University

"Minimum cost path planning in the random traversability space," A. Meystel, Drexel University

"Algorithms to reconstruct a convex set from sample points," Marc Moore, Ecole Polytechnique Montreal and McGill University, Y. Lemay, Bell Canada, and S. Archambault, Ecole Polytechnique Montreal

"On the geometric probability of discrete lines and circular arcs approximating arbitrary object boundaries," Chang Y. Choo, Worchester Polytechnic Institute

"Application of orthogonalization procedures to fitting tree-structured models," Cynthia O. Siu, The Johns Hopkins University

"Evaluation of functions over lattices," Michael Conlon, University of Florida

SATURDAY, APRIL 23, 1988

10:45 a.m. - 12:00 noon — **Room 6**

Special Invited Lecture III, Chaired by: Sally Howe, National Bureau of Standards

"Visualizing high dimensional spaces," Thomas Banchoff, Brown University

10:45 a.m. - 12:45 p.m. — **Room 5**

Entropy Methods, Chaired by: Raoul LePage, Michigan State University

"Introduction to relative entropy methods," John Shore, Entropic Processing Corporation

"Structural covariance matrices and 2-dimensional spectra," John Burg, Entropic Processing Corporation

"Matrix completion and determinants," Charlie Johnson, College of William and Mary

10:45 a.m. - 12:45 p.m. — **Room 3**

Contributed Papers: Information Systems, Databases and Statistics, Chaired by: Robert Teitel, Teitel Data Services

"Information systems and statistics," Nancy Flournoy, National Science Foundation

"Is there a need for a statistical knowledge base?" Z. Chen, Louisiana State University

"An alternate methodology for subject database planning," Craig W. Slinkman, Henry D. Crockett, and Mark Eakin, University of Texas at Arlington

"A sensitivity analysis of the Herfindal-Hirschman Index," James R. Knaub, Jr., U. S. Department of Energy

"Statistical methods for document retrieval and browsing," Jan Pedersen, Xerox PARC and John Tukey and P. K. Halvorsen

10:45 a.m. - 12:45 p.m. — **Room 2**

Contributed Papers: Parallel Computing, Chaired by: Joseph Brandenburg, INTEL Scientific Computers

"Programming the BBN butterfly parallel processor," Pierre duPont, BBN Advanced Computers

"A tool to generate parallel FORTRAN code for the Intel iPSC/2 hypercube," Carlos Gonzalez, J. Chen and J. Sarma, George Mason University

"All-subsets regression on a hypercube multiprocessor," Peter Wollan, Michigan Technological University

"Multiply twisted N-cubes for multiprocessor parallel computers," T.H. Shiau, University of Missouri, Columbia

"Markov chains arising in collective computation networks with additive noise," R.H. Baran, Naval Surface Warfare Center

SATURDAY, APRIL 23, 1988

10:45 a.m. - 12:45 p.m. Room D

Contributed Papers: Density and Function Estimation, Chaired by: Celesta Ball, George Mason University

"The L_1 asymptotically optimal kernel estimate," Luc Devroye, McGill University

"Derivative estimation by polynomial-trigonometric regression," Paul Speckman, University of Missouri, Columbia and R.L. Eubank, Southern Methodist University

"A pooled error density estimate for the bootstrap," Walter Liggett, National Bureau of Standards

"Efficient algorithms for smoothing spline estimation of functions with or without discontinuities," Jyh-Jen Horng Shiau, University of Missouri, Columbia

"On the convergence of variable bandwidth kernel estimators of a density function," Ting Yang, University of Cincinnati

10:45 a.m. - 12:45 p.m. Room 1

Contributed Papers: Statistical Methods IV, Chaired by: LeRoy A. Franklin, Indiana State University

"Stochastic test statistics," P. Warwick Millar, University of California, Berkeley

"It's time to stop!," Hubert Lilliefors, George Washington University

"The effects of heavy tailed distributions on the two sided k-sample Smirnov test," Henry D. Crockett and M. M. Whiteside, University of Texas at Arlington

"Performance of several one sample procedures," David Turner, Utah State University

"Exact power calculation for the chi-square test of two proportions," Carl E. Pierchala, Food and Drug Administration

Abstracts

Abstracts are arranged in alphabetical order of the last name of the first author. The first author may not correspond to the presenter of the paper. Thus in looking up an abstract for a paper, it may be worthwhile to search under co-authors. In any case, the abstracts are referenced in the author index and may be located by use of the index .

A

A Probabilistic Approach to Range Image Description

E. Al-Hujazi
Wayne State University

and

A. K. Sood
George Mason University

In this paper we present an approach for describing range images based on the H (Mean Curvature) and K (Gaussian Curvature) parameters. Range images are unique in that they directly approximate the physical surfaces of a real world 3-D scene. H and K are defined from the fundamental theorems of differential geometry, and provide visible, invariant pixel labels that can be used to characterize the scene. The sign of H and K can be used to classify each pixel into one of eight possible surface types. Due to sensitivity of these curvature parameters to noise, the computed HK-sign map does not directly identify surfaces in the range image. In this paper a probabilistic approach for the segmentation of the HK-sign map is suggested. The image is modeled as a Markov random field on a finite lattice. The prior knowledge about the solution is expressed in the form of a Gibbs probability distribution. This approach allows the integration of the output of a number of modules in an efficient way. Due to the computational complexity of this approach, a sub-optimal algorithm using dynamic programming has been developed. The performance of the proposed techniques on a number of range images will be presented.

Image Analysis of a Turbulent Object Using Fractal Parameters

Amar Ait-Kheddache
North Carolina State University
Electrical and Computer Engineering Department
Campus Box 7911
Raleigh, NC 27695-7911

The objective of this paper is threefold. First, it describes the use of image processing techniques for recording and measuring information about pollutant dispersion (smoke). Visual images of the smoke plume dispersion are used to develop techniques for describing wake processes. Second, a new model based on fractal concepts is developed to analyze smoke data. The concept of fractals is introduced for the purpose of giving some qualitative and quantitative interpretation to the transient flows of the pollutant. The fractals display interesting dynamics and provide an environment for modeling complex natural phenomena. Third, a theoretical justification and mathematical methods are developed for making the concept useful in practice. We have chosen two fractal parameters, the horizontal fractal parameter and the vertical fractal parameter to characterize the image data. These parameters are computed only for the very active regions (turbulent regions) of the phenomenon (smoke) and they are nonconservative properties. Analysis and testing of the technique have determined information about which features can be extracted from the image sequences (spatio-temporal characteristic, concentration, velocity...). Some statistical interpretation which support the results are reported. The limitations of the techniques are also addressed. In summary, the phenomenon itself, the experimental study and the achieved results using fractals constitute the novelty of the work.

Smoothing Data with Correlated Errors

Naomi Altman

Cornell University

Suppose the dependent variable y is observed with error at a set of design points x on an interval, and that the mean of y is assumed to be a smooth function of x. Linear nearest neighbors, kernel regression estimators, and smoothing splines are all examples of techniques for estimating the mean function which depend on a single smoothing parameter, λ, and are linear functions of the data when λ is fixed.

When the error process is weakly continuous, ther is a non-zero lower bound on the variance of linear estimators of the mean as the sample size increases on a fixed interval. So the estimators cannot converge in any sense to a deterministic function, as they do when the errors are independent.

The standard techniques for selecting smoothing parameters, such a cross-validation and generalized cross-validation, perform very badly when the errors are correlated. If the sum of the correlations from zero to infinity is negative, the techniques favor oversmoothing; if the sum is positive, the techniques favor undersmoothing. However, the selection criteria can be adjusted to incorporate the known effects of the correlations or the residuals on which the criteria are based can be transformed to eliminate the effects of correlations.

Estimates of the correlation function based on residuals from a preliminary smooth are shown to be very biased. Oversmoothing leads to estimates of correlation which are too large, whiler undersmoothing leads to estimates which are too small. This leads to a negative feed-back effect which makes iterative techniques inadvisable.

In simulation, the standard selection criteria are shown to behave as predicted by the theory. The corrected criteria are shown to be very effective when the correlation function is known. Although the estimates of correlation based on the data are poor, they are shown to be sufficient for correcting the selection criteria, particuloarly if the signal to noise ratio is small.

LINEAR PREDICTION OF FAILURE TIMES OF A REPAIRABLE SYSTEM

M. Ahsanullah

Rider College Lawrenceville, New Jersey 08648-3099, U. S. A.

ABSTRACT

Suppose we consider a repairable system in which a failed component is replaced immediately by a component of equal age. On replacement of the component, the system becomes operational and we assume the repairing time of the component is negligible. We assume the survival times of the components are independent and identically distributed.

Let us denote by X_o, X_1, X_2, ... the failure times of the system where $X_o = 0$. The time between failures $U_n = X_n - X_{n-1}$ $n \geq$ are non negative random variables. Let $F(t) = \Pr(U, \leq t)$, for $t \geq 0$ and $\bar{F}(t) = 1 - F(t)$.

We assume that F(t) has a density f(t) with F(0) = 0 and $r(t) = f(t)/\bar{F}(t)$, for $\bar{F}(t) > 0$. The function r (t) is called hazard rate and $R(t) = \int_0^t r(u)\,du$ is called the cumulative hazard rate. The hazard rate of the system after repair is assumed to be the same as before. Let $F_n(t) = \Pr(X_n \leq t)$ and $f_n(t) = F_n'(t)$. Then

$$1-F_n(x) = \begin{cases} \bar{F}(x) \text{ if } n = 1 \\ \bar{F}(x) + \bar{F}(x), \text{ if } n=2 \end{cases}$$

and in general,

$$1 - F_n(x) = \bar{F}(x) \sum_{i=o}^{n-1} (R(x))\ i/i!$$

$1-F_n(x)$ can be interpreted as the survival time to the n^{th} failure of the system given that a failed component is replaced by one of equal age and the repair time is negligible. The density $f_n(x)$ of X_n can be written as $f_n(x) = f(x). \dfrac{(R(x))^{n-1}}{(n-1)!}$, $n \geq 1$.

Some distributional properties of the n^{th} survival times are discussed when F has different life distributions. Various predictions of the sth failure time x_s (s>n), based on the first n and as well as on some selected failure times are obtained. Their expected costs with respect to different cost functions and a replacement Model, where the system is replaced at a certain failure or failure time, are computed.

Fusion and Propagation in Graphical Belief Models

Russell Almond
Harvard University

ABSTRACT

Graphical models are a clear and concise way of describingprobabilistic dependencies among many variables. Only relationships between variables which share a common hyperedge are modeled, considerably simplifying both the modeling and the computational tasks. The latter represents considerable savings, as the direct approach to calculating marginal relationships from the components of a graphical model is computationally expensive, requiring 2^2 operations for n binary variables. Graphical models have lately been studied by Pearl [1986a,1986b], Moussouris[1974], and Lauritzen and Spiegelhalter[1987] in the Bayesian case, and Kong[1986a], Shafer, Shenoy, and Mellouli [1986], and Shenoy and Shafer[1986] in the Belief Function case.

Belief functions are a generalization of probability measures that allow ways to express total ignorance, Bayesian prior probability distributions, conditional probability distributions (likelihoods), logical relationships (production rules) and observations. All these diverse types of knowledge can be combined with a uniform fusion rule, the direct sum operator. Simple procedures can restrict belief functions to a smaller frame and extend them to a larger frame without adding additional information. The theory of belief functions is developed by Dempster[1967], Shafer[1976,1982], and Kong[1986a].

By a simple procedure given here and in Kong[1986b], we can transform the model hypergraph into a *{tree of closures}*. I present apròpagation algorithm from Dempster and Kong [1986] for finding marginal belief functions from a tree of closures. Each node of the tree of closures is a "chunk" of the original problem; each chunk can be computed independently of all other chunks except its neighbors. Every node in the tree passes to each of its neighbors a message (expressed as a belief function) that consists of the local information fused with all of the information that has propagated through the other branches of the tree. Using this propagation algorithm along with the fusion algorithm given by the direct sum operator, we can easily compute marginal beliefs, with substantially less computational cost than the direct approach. I have translated this mathematical formalisim into a computer program and dicuss some examples computed using this procedure.

Key Concepts: *Graphical Models, Belief Functions, Bayesian Models, Fusion and Propagation, Probability in Expert Systems, Triangulated Graphs.*

Abstract Bibliography

Dempster[1967], "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics*, Vol. 38, 325-339.

Dempster and Kong[1986], "Uncertain Evidence and Artificial Analysis," technical report S-108, Harvard University, Department of Statistics.

Kong[1986a], "Multivariate Belief Functions and Graphical Models," Ph.D. thesis, technical report S-107, Harvard University, Department of Statistics.

Kong[1986b], "Construction of a Tree of Cliques from a Triangulated Graph," technical report, Harvard University, Department of Statistics.

Lauritzen and Spiegelhalter[1987], "Fast Manipulation of Probabilities with Local Representations--With Applications to Expert Systems," technical report R-87-7, Aalborg University, Institute of Electronic Systems, Denmark.

Moussouris[1974], "Gibbs and Markov Random Systems with Constraints," *Journal of Statistical Physics*, Vol. 10, pp. 11-33.

Pearl[1986a], "Fusion, Propagation, and Structuring in Belief Networks," *Artificial Intelligence*, Vol. 29, pp. 241-288.

Pearl[1986b], "Markov and Bayes Networks: A Comparison of Two Graphical Representations of Probabilistic Knowledge," technical report R-46, University of California at Los Angles, Computer Science Department.

Shafer[1976], *A Mathematical Theory of Evidence*, Princeton University Press.

Shafer[1982], "Belief Functions and Parametric Models," *Journal of the Royal Statistical Society, Series B*, Vol. 44, pp. 322-352.

Shafer, Shenoy, and Mellouli[1986], "Propagating Belief Functions in Qualitative Markov Trees," working paper no. 196, University of Kansas, School of Business.

Shenoy and Shafer[1986], "Propagating Belief Functions with Local Computations," *IEEE Expert*, Vol. 1 No. 3, pp. 43-52.

A BAYESIAN MODEL OF INFORMATION COMBINATION FROM NOISY SENSORS

G. Anandalingam

Department of Systems
University of Pennsylvania
Philadelphia, PA 19104-6315

A paper to be presented at

The 20th Symposium on the
Interface of Computing Science and Statistics
Reston, Virginia. April 21-23, 1988

ABSTRACT

An important thrust of research in artificial intelligence (AI) has been the use of multiple sensors (or experts) for information processing. The work that falls into this category is often called "Distributed AI". Researchers worry about the placement of these sensors (choice of experts), and ways to combine the distributed corpus of knowledge. Parallel, and somewhat preceding these research thrusts, a number of statisticians have been working in the area of combining statistical data, and management scientists have been working on the combination of time-series forecasts. The main problem in all these studies has been the extraction of weights for the individual information sources.

In this paper, we use a Bayesian approach to combine information from distributed sensors. We extend and generalize previous Bayesian analyses to incorporate noisy information, and lagged sensor responses. In order to do the latter, we show the connection between the generalized Bayesian model, and Kalman Filtering in dynamic systems analysis. In all cases, the combined information is shown to be unbiased (i.e. unaffected by measurement errors in the sensors) and efficient.

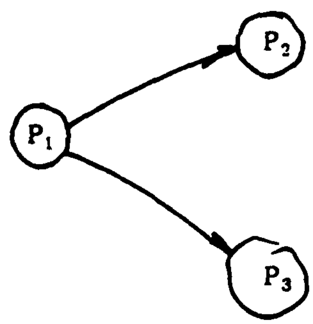
We also examine the case where the sensor error structures are unknown to the information processor. We set up a Bayesian procedure to learn about the sensors, and to combine information recursively. The learning feature is novel for the statistical literature on information combination, but is well in the spirit of artificial intelligence research.

A stochastic extension of Petri Net graph theory

Lisa M. Anneberg
Wayne State University

A Petri net is a bipartite graph, and is heavily utilized for modelling computer hardware and software (among other items). The two nodes (arcs and places) will each have an associated probability (of correct operation) and two time values (average time waiting and time elapsed during function). The probabilities associated with both places and transition can give both the overall reliabilities of all paths, and each place/transition pair reliability.

A small example net will serve to illustrate this idea, with the asssociated place transition matrix:



$$P \times T = \begin{bmatrix} p(P_1)p(t_1) & p(P_1)p(t_2) \\ p(P_2)p(t_1) & p(P_2)p(t_2) \\ p(P_3)p(t_1) & p(P_3)p(t_2) \end{bmatrix}$$

$$= \begin{bmatrix} 0.04 & 0.36 \\ 0.06 & 0.00 \\ 0.00 & 0.54 \end{bmatrix},$$

where $p(P_1) = 0.4$, $p(P_2) = 0.6$, $p(P_3) = 0.7$, $p(t_1) = 0.1$, and $p(t_2) = 0.9$.

One cannot, however, arrive at total path reliabilities via this matrix because interior arc/place probabilities will be counted twice. For particular place/transition or transition/place pairs, this matrix shows the proper reliabilities. Each set of reliabilities is useful. The place x transition matrix can identify the critical place/transition paris that may be pulling a corresponding overall path reliability quite lower.

Times associated with place/transition pairs can be represented in this fashion (addition instead of multiplication is used, of course). Again, this identifies critical pairs, but cannot be utilized to arrive at an overall time unless the double counted interior nodes are accounted for.

A short technical paper will be presented elaborating on these points.

Identification of Closed Figures

Jeff Banfield

Department of Mathematical Sciences
Montana State University

Adrian Raftery

Department of Statistics
University of Washington

ABSTRACT: A recurring problem in image processing is the recognition and representation of closed figures. A technique to solve this problem, incorporating several innovative new ideas, is illustrated by locating ice floes in a LANDSAT image. Using standard image processing techniques, the image pixels are classified as ice or water and the edge pixels (those which define the border between ice and water) are identified. The ice floes are then eroded by the computer to simulate melting the ice. The locations of those edge pixels which outline a given floe are propagated into the interior of the floe as it melts. This results an initial clustering of the edge pixels which belong to the larger floes and the elimination of edge pixels from noise and floes smaller than a specified size. A new clustering criteria, based upon principal curves and maximum likelihood estimation, is used for the final identification and representation of the floes.

MARKOV CHAINS ARISING IN COLLECTIVE COMPUTATION NETWORKS WITH ADDITIVE NOISE

R. H. Baran
Naval Surface Warfare Center, White Oak (code U23, rm. 2-250),
Silver Spring, MD 20903-5000

Recent progress in the modelling of connectionist ("neural") networks gives rise to the expectation that future computing systems will employ coprocessors in which large numbers of memoryless, nonlinear processing units interact through plastic connections. Hopfield has drawn attention to a class of networks, defined by symmetric interconnections and processing units with binary-valued outputs, which can compute good (suboptimal) solutions to difficult constrained optimization and decision problems. These collective computation networks (CCNs) converge rapidly to stable states which correspond to local minima of the computational energy, a bilinear functional of the network state vector. The CCN can be freed from local minima by the addition of noise to the input of each processing unit (or "neuron"). The network state then takes a random walk on a lattice of 2^N points, where N is the number of "neurons". Ackley, Hinton, and Sejnowski have suggested that the long term evolution of the state (K) follows a Boltzmann distribution,

$$\Pr(K = k) = \frac{\exp(-E_k/T)}{\sum_k \exp(-E_k/T)}, \quad k = 0, 1, \ldots, 2^N-1,$$

where E_k is the computational energy of the k-th state and T is the "temperature".

This paper uses a simple, explicit algorithm to study the behavior of "Boltzmann machines" having various configurations and noise distributions. The two-neuron network is analyzed in detail to obtain an expression for the effective temperature. That the result generalizes to larger networks is verified by Monte Carlo calculations in which the randomly sampled state exhibits a distribution that is statistically close to the theoretical.

Statistical Learning Networks: A Unifying View

Andrew R. Barron
Statistics and Electrical and Computer Engineering Departments
University of Illinois
and
Roger L. Barron
Barron Associates, Inc.
Stanardsville, Virginia

We trace the history of artificial neural network models from the viewpoint of 25 years of involvement in the application of these models to curve-fitting problems (involving regression, prediction, classification, or guidance and control) in specific projects for government and industry. Although originally some of these network models were derived from analogies to neurophysiological systems, the driving force in the development has been practical empirical modeling problems. The characteristic shared by each of these methods is that estimates of functions of many variables are obtained by the mathematical composition (interconnection) of many simple relationships. It is therefore suggested that the name *statistical learning network* rather than *neural network* more accurately conveys the nature and purpose of these models.

It is recounted how the advancement of learning network methodologies has depended on statistical developments (nonparametric smoothing, model selection criteria, asymptotic theory), information-theoretic developments (universal data compression, complexity minimization), and computational developments (efficient search techniques for multimodal surfaces) as well as developments in approximation theory (What classes of functions are approximated by functions expressed by networks?). We describe the surprising similarities as well as the differences between learning network models such as fixed *polynomial networks* (devised by Snyder, Barron et.al. 1964 and described in Gilstrap 1971), *adaptively synthesized networks* (developed by Mucciardi 1970, Ivakhnenko 1971, and Barron et.al. 1984), *projection pursuit* (Friedman and Tukey 1974, Friedman and Stuetzle 1981), and classifiers trained by *back-propagation* (Rumelhart, Hinton and Williams 1986). A flexible system of computer programs is being developed to implement these and many other learning network models according to user specified attributes.

Some approximation theory questions concerning functions represented by networks are resolved. A four layer polynomial network of depth 2m+1 and fixed connectivity can uniformly well approximate any continuous function of m variables on a compact set. Similarly for projection pursuit, it is known that the theoretical (non-sampling) version approximates any L2 function of m variables (Jones 1987). A fundamental statistical question remains: Do estimated networks converge to the unknown function with high probability as the sample size increases without bound? No consistency or rate of convergence results are yet available for any of these learning network estimators. Recent results (Barron 1987) concerning Bayes estimators for nonparametric smoothing and complexity minimization show promise for helping resolve some of these consistency questions.

References

A. R. Barron, The Exponential Convergence of Posterior Distributions with Implications for Bayes Estimators of Density Functions, submitted to the *Annals of Statistics*, 1987.

R. L. Barron, A.N. Mucciardi, F.J. Cook, J.N. Craig, and A.R. Barron, Adaptive Learning Networks: Development and Application in the United States of Algorithms Related to GMDH, in *Self-Organizing Methods in Modeling*, S.J. Farlow, editor, Marcel Dekker, New York, 1984.

J. H. Friedman and J. W. Tukey, A Projection Pursuit Algorithm for Exploratory Data Analysis, *IEEE Transactions on Computers*, Vol.23, pp. 881-889, 1974.

B

J. H. Friedman and W. Stuetzle, Projection Pursuit Regression, *Journal American Statistical Association*, Vol.76, 817-823, 1981. L. O. Gilstrap, Jr., Keys to Developing Machines with High-Level Artificial Intelligence, *Proceedings ASME Design Engineering Conference*, Apr. 1971, ASME Paper 71-DE-21.

A. G. Ivakhnenko, Polynomial Theory of Complex Systems, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.1, pp.364-378, 1971.

L. K. Jones, On a Conjecture of Huber Concerning the Convergence of Projection Pursuit Regression, *Annals of Statistics*, Vol.15, pp.880-882, 1987.

A. N. Mucciardi, Neuromime Nets as the Basis for the Predictive Component of Robot Brains, *Cybernetics, Artificial Intelligence, and Ecology*, H.W. Robinson and D.E. Knight (Eds.), Spartan Books, 1972, pp.159-193. (Presented at 4th Annual Symposium American Society of Cybernetics, Oct. 1970.)

D. E. Rumelhart, G.E. Hinton, and R.J. Williams, Learning Internal Representations by Error Propagation, in D.E. Rumelhart and J.L. McClelland, *Parallel Distributed Processing*, op.cit. 1986.

R. F. Snyder, R.L. Barron, et al., *Advanced Computer Concepts for Intercept Prediction*, Vol.I: *Conditioning of Parallel Networks for High-Speed Prediction of Re-entry Trajectories*, Adaptronics, Inc. Final Technical Report, Army Nike-X Project Office, Redstone Arsenal, Ala., Nov. 1964.

Interior Point Methods for Linear Programming Problems

Paul T. Boggs

Scientific Computing Division
National Bureau of Standards
Gaithersburg, MD 20899

ABSTRACT

The method of centers was first proposed by Huard for convex nonlinear optimisation problems. A version of the method was shown to be a polynomial algorithm for the linear programming problem. Moreover, the order of the polynomial is the same as for Karmarkar's method. In this talk, the basic method as applied to linear programmming is described and a continuous version derived. The continuous version yields trajectories from any feasible point in the polytope to the solution. The properties, including the deficiencies, of these trajectories are discussed. A modification that overcomes the difficulties is proposed and analyzed. Finally, an algorithm based on these results is given and some preliminary numerical results are presented.

On Some Graphical Representations of Multivariate Data

Masood Bolorforoush
Edward J. Wegman

George Mason University

The paper presents an implementation of some multivariate graphical techniques written in PASCAL and developed for the IBM-RT. We have a basic implementation of the parallel coordinate representation together with some enhancements including brushing, windowing, zooming, and transformations including Box-Cox and standardization. Also included in our package are scatter plot diagrams which may be linked in split screen to parallel coordinate diagrams. Some related techniques which we call color histograms and relative slope plots are also implemented.

A MONTE CARLO ASSESSMENT OF CROSS-VALIDATION AND THE C_p CRITERION FOR MODEL SELECTION IN MULTIPLE LINEAR REGRESSION

Robert M. Boudreau

Dept of Math. Sciences, Virginia Commonwealth University

For selecting variables or model building in the multiple linear regression situation, Mallows C_p criterion is relevant when the regressors are considered fixed. When the regressors are random, then cross-validation is more appropriate. Both these methods are often justified on the grounds that they estimate the unobservable conditional prediction mean squared error (PMSE) when predicting new observations using the current training data set to estimate the parameters. In the fixed case, a theoretical result is given showing that the C_p for a given model is in fact uncorrelated with the training set PMSE. In the case of random regressors, results of a simulation experiment, with some related theory, give evidence that cross-validation (counter to intuition) is also uncorrelated, or at most weakly correlated, with the PMSE for that data set.

COMPUTATIONAL EXPERIENCE WITH THE GENERALIZED SIMULATED ANNEALING ALGORITHM

Daniel G. Brooks William A. Verdini

Arizona State University

Computational results using the generalized simulated annealing algorithm are presented. The algorithm is used on a number of well-known test problems and solution results are compared to those of other stochastic optimization procedures. The sensitivity of the rate of convergence to changes in several algorithm parameters is presented.

B

AN ALGORITHM TO IDENTIFY CHANGES IN HORMONE PATTERNS

Morton B. Brown
Department of Biostatistics
The University of Michigan
Ann Arbor, MI 48109-2029

Fred J. Karsch and Benoit Malpaux
Consortium for Research in
Developmental and Reproductive Biology
The University of Michigan

Many hormones are secreted into the blood in a pulsatile manner: i.e., in high concentrations at 'random' intervals. To study hormone levels, researchers assay its level in the blood at regularly spaced intervals. The statistical problem is to differentiate between changes in stage (level of the hormone) and observations influenced by a 'random' pulse ('noise'). An algorithm is described that uses regression-like statistics computed after deleting the most 'extreme' observation combined with a moving variable-length window to identify rises and declines in hormone level. The deletion of the most 'extreme' observation and the use of a variable-length window facilitates the exclusion of 'noisy' values from the determination of the stage of the hormone.

Keywords: hormone levels
circadian and annual rhythms
pattern analysis
regression

BOOTSTRAPPING REGRESSION STRATEGIES

by: David Brownstone
School of Social Sciences
University of California
Irvine, California 92717
Tel: 714-856-6231
Bitnet: "DBROWNST@UCI"

Applied statisticians rarely estimate multiple regression models with a single estimator; they follow complex estimation strategies using many related models, estimators, and diagnostic statistics. Although it is known that the use of these strategies can create large biases in standard dispersion measures from the final estimates, there has been very little work on quantifying these biases due to the analytic intractability of the problem. This paper demonstrates the feasibility of using bootstrap techniques to estimate the sampling distribution of regression estimation strategies. A number of Monte Carlo experiments are performed using Ordinary Least Squares on a small 5 variable regression model. We consider simple strategies like deleting all variables corresponding to nuisance parameters with t-statistics less than 2 and then reestimating the model. These experiments verify that common simple estimation strategies can create large biases in standard dispersion estimators, and the magnitude of these biases depends on both the true model design and estimation strategy.

The bootstrap methodology can be applied to more realistic, complex strategies and estimators. We demonstrate this with experiments where outliers are removed before the models are reestimated. Removing outliers can either increase or decrease dispersion estimator bias depending on whether outliers are unusual draws from a well behaved distribution or "normal" draws from a fat-tailed or contaminated distribution.

The computations for this paper were performed on PC and PC/AT computers using the GAUSS programming language. On more powerful workstations, it would be feasible to bootstrap more complex strategies found in expert regression systems such as AT&T Bell Laboratory's REX system. The results of the Monte Carlo experiments performed here strongly suggest that the biases in parameter dispersion estimators increase with the complexity of the estimation algorithm. The bootstrap techniques presented here are the only practical way to generate consistent estimates of parameter dispersion for complex regression estimation strategies. Bootstrapping could also be incorporated into expert systems for multiple regression models. This would greatly improve the reliability of the dispersion estimates for the final model produced by these systems.

Noise Appreciation: Analyzing Residuals Using RS/Explore

David A. Burn
Fanny L. O'Brien

BBN Software Products Corporation
10 Fawcett Street, Cambridge, Massachusetts

The RS/Explore software is a statistical advisory environment for performing analysis of general linear models. One goal of data analysis is to find a "model" that adequately describes the variation in the data. Residual analysis is an invaluable tool in selecting and validating a model. We will examine how RS/Explore provides a convenient access to traditional and innovative graphical displays useful in residual analysis.

ROBUSTNESS OF WEIGHTED ESTIMATORS OF LOCATION: A SMALL-SAMPLE STUDY

Gregory Campbell and Richard I. Shrager
Division of Computer Research and Technology
National Institutes of Health

The problem of location estimation is considered in the context of known as well as misspecified weights. For the one-sample problem, the studied estimators include weighted analogs of the mean, the median, the median of Walsh averages, Huber M-estimators and a computer-intensive procedure which minimizes the weighted sum of the absolute deviations. For estimators which employ a weighted median, interpolation to improve performance is considered. The estimators are evaluated by computer simulation with respect to robustness to weight misspecification as well as robustness to outliers. The Kantorovich inequality provides additional insight concerning the small-sample efficiency of estimators with misspecified weights.

Neural Petri Nets

N. H. Chamas
Wayne State University

It is shown that Petri nets have been evolved into a powerful tool for analyzing asychronous concurrent systems. But the task complexity in digital computers is still high in emulating natural information processing that humans can routinely handle. Billions of operations in a sequential machine that may take hours or days may take only seconds for the human brain. This work clarifies the similarity between the neural cell and a Petri net. The similarity will be illustrated by an example. Figure 1 is a typical neural cell while Figure 2 is a typical Nenura Petri Net (NPN).

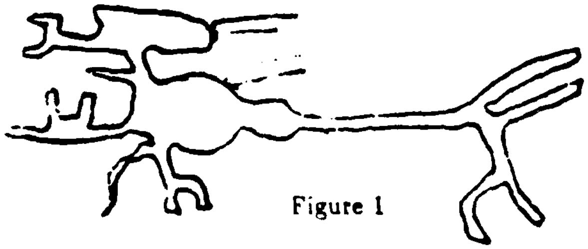


Figure 1

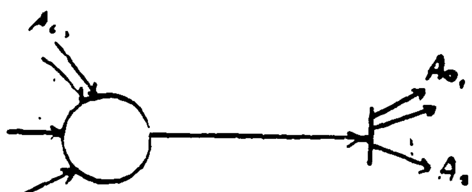


Figure 2

The places and the transitions in NPN have some properties different from the properties and transitions in PN. The main difference is that the place in NPN has onle one output and many inputs, and the transition in NPN has one input and many outputs. These properties make the NPN place similar to the soma in the neural cell, the transition similar to the hillock, and the arcs similar to the axon terminals. New rules on concurrency and computation will be illustrated and new approaches will be proposed.

APPROACHES FOR EMPIRICAL BAYES CONFIDENCE INTERVALS WITH APPLICATION TO EXPONENTIAL SCALE PARAMETERS

Bradley P. Carlin and Alan E. Gelfand
University of Connecticut

ABSTRACT

Parametric empirical Bayes methods of point estimation date to the landmark paper of James and Stein (1961). Interval estimation through parametric empirical Bayes techniques has a somewhat shorter history, which is summarized in the recent paper of Laird and Louis (1987). In the i.i.d. exchangeable case, one obtains a "naive" EB confidence interval by simply taking appropriate percentiles of the estimated posterior distribution of the parameter, where the estimation of the prior parameters ("hyperparameters") is accomplished through the marginal distribution of the data. Unfortunately, these "naive" intervals tend to be too short, since they fail to account for the variability in the estimation of the hyperparameters. That is, they don't attain the desired coverage probability, both in the classical sense and in the "EB" sense defined in Morris (1983).

In this paper we consider two methods for developing EB intervals which attempt to correct this deficiency in the naive intervals. The first is a "bias corrected naive" method inspired by Efron (1987). Simply put, this method adjusts the naive intervals using tail areas determined by the parametric structure of the model and the data. In certain cases these adjusted tail areas can be found using only a simple rootfinding algorithm; in more complicated settings one likely needs to bootstrap, as suggested by Efron. The second method addresses tranformations of the bootstrap observations to match a specified hyperprior Bayes solution. In this context we clarify the nature of Laird and Louis' Type III parametric bootstrap.

To compare the four types of EB intervals (naive, bias-corrected naive, Laird and Louis, and hyperprior matched) we compute expected "true" tail areas and "true" interval lengths (as developed in Laird and Louis), as well as simulated coverage probabilities and interval lengths. This is done illustratively in the context of confidence intervals for a vector of exponential scale parameters.

References

1. Efron, B. (1987), Comment on "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," by N.M. Laird and T.A. Louis, *Journal of the American Statistical Association*, 82, 754.

2. James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 361-379.

3. Laird, N.M., and Louis, T.A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," *Journal of the American Statistical Association*, 82, 739-750.

4. Morris, C.N. (1983), "Parametric Empirical Bayes Confidence Intervals," in *Scientific Inference, Data Analysis, and Robustness*, New York: Academic Press, pp. 25-50.

C

INFERENCE TECHNIQUES FOR A CLASS OF EXPONENTIAL TIME SERIES

V. Chandrasekar and P.J. Brockwell

Colorado State University

This research has been motivated by the need to study meteorological radar signals. The power received by a radar backscattered from randomly position and moving targets is a time series with exponential margional distributions. Moreover the signals are observed at two polarizations states of the transmitted wave are correlated. The observations are made alternating between the polarization states and as a result we have missing samples at any polarization.

In this paper we discuss the inference problems associated with the above described radar signals. The radar signals are obtained from a multivariate complex guassian series. We discuss different inference schemes in the context of applicability in real time implementation for radar systems. Time series data collected using radar observations of rainfall are used to compare against model results.

Paper for Interface88

IS THERE A NEED FOR STATISTICAL KNOWLEDGE BASE?

(Abstract)

Z. Chen,

P. O. Box 22236, LSU,

Baton Rouge, LA 70893

Statistical knowledge base! This means to explicitly store statistical knowledge in the knowledge base. Although statistics has long been involved in abductive reasoning (since MYCIN), the involvement of statistics in knowledge engineering is very limited, and it is almost around the use of Bayes's theorem. The coming of statistical knowledge base will make statistics the first order citizen in the research of knowledge engineering. But is there a need for such a new concept?

In this paper we argue that this kind of need does exist. First of all, *statistical knowledge* exists at its own right, it plays not only a role of measurement. Secondly, making statistics as the first order citizen means the whole set of matured statistic methods (eg. multivariate analysis) can be used in knowledge engineering. Finally, the method of abductive reasoning itself can be enriched: for instance, searching in abduction will no longer be restricted to a bottom-up manner.

In the rest of this paper we discuss the possible interface of statistical knowledge base and current existing statistics software. We also compare the similarity and difference between statistical database and statistical knowledge base.

EFFICIENT DATA SENSITIVITY COMPUTATION FOR MAXIMUM LIKELIHOOD ESTIMATION

Daniel C. Chin and James C. Spall

The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, MD 20707

Abstract

A computational procedure and numerical results are presented for studying the effect of outliers or other anomalous data. This procedure is based on a first order approximation relying on the implicit function theorem, and involves matrix operations and tensor (Kronecker) algebra. The approximation yields a closed form expression; in contrast, the calculation of the MLE depends on iterative numerical methods such as Newton-Raphson, steepest descent, or scoring. The approximation is generally much more efficient than a straightforward computation of the MLE via such numerical methods. We will present the results of a numerical study that illustrate the procedure on a multivariate signal-plus-noise problem with non-identically distributed noise. Such signal-plus-noise estimation problems arise in many settings (e.g., Kalman filter model estimation, dose response curve estimation, etc.). In the numerical study we compared this procedure with the scoring method for finding MLES. In a moderate size problem we found that the procedure was more than 25 times faster; greater computational savings would be expected in a larger dimensional problem.

Keywords and phrases: Computational Stochastic Modeling, MLE approximation, numerical methods, simulation study, outliers, signal-plus-noise models.

On the Geometric Probability of Discrete Lines and Circular Arcs Approximating Arbitrary Object Boundaries

Chang Y. Choo
Department of Electrical Engineering
Worcester Polytechnic Institute
Worcester, MA 01609

Grid-based line data representation such as chain codes and polycurve codes is an efficient scheme used for representing arbitrary object boundaries in the areas of image processing and pattern recognition. Grid-based schemes of representing object boundaries consist of three processes. First, a square grid of proper size is overlaid onto the boundaries. Second, connected straight-line and circular-arc segments, each of which is predefined with respect to grid points, are searched that best fit all the grid intersection points. Finally, according to predetermined rules, each segment is mapped into an integer and stored in a computer. The number of discrete lines and circular-arc segments used as approximators increases rapidly as the size of "quantization window", in which one curve fitting is done, increases.

This paper addresses the issue of calculating the probability of the line and circular-arc segments based on a model of random line drawing within a quantization window. The model assumes that the original line drawing inside a quantization window is a random circular arc. According to a quantization algorithm, the probability that each line or circular-arc segment will be used for approximating the random line drawing is calculated. The analytical results are verified by various experiments involving real object boundaries and map contour lines. The results of this paper may be used to design variable-length codes such as Huffman codes.

C

Computing Extended Maximum Likelihood Estimates in Generalized Linear Models

by

Douglas B. Clarkson
IMSL, Inc.

Robert I. Jennrich
University of California, Los Angeles

Abstract

Concern here is with computing the "extended maximum likelihood" estimates of Haberman (1974) in which one or more parameter estimates is infinite at the supremum of the likelihood. Theorems justifying the computation of these estimates are presented in a general context and efficient algorithms for detecting and computing such estimates in the context of generalized linear models are given. Examples illustrating the use of these algorithms are presented.

Evaluation of Functions over Lattices

Michael Conlon
Department of Statistics, University of Florida, Gainesville, Fl

Consider the problem of evaluating the sum of a function of two arguments over a subset of a lattice of argument values. A new recursive algorithm has been developed which performs these evaluations at considerable savings when portions of the lattice can be identified as contributing little to the overall sum. The algorithm takes full advantage of adjacency relationships. Each function evaluation after the first can be performed using prior knowledge of an adjacent function value on the lattice. The algorithm has been applied to computing functionals of estimators for comparative binomial experiments. Exact evaluation of expected value, variance, and other functionals can be computed from basic principles using the new algorithm in one order of magnitude less time than performing a simulation.

Extracting Records from New Jersey's Multiple Cause of Death Files

Giles Crane

New Jersey Department of Health

A simple microcomputer system has been developed using of-the-shelf components which permits local access in an acceptable time frame to several years of New Jersey multiple cause of death data assembled and distributed by the National Center for Health Statistics. The system includes hardware and software and illustrates a trade-off between speed and specificity of access to approximately 70,000 records per calendar year. Applications to the epidemiology of drowning and sickle cell anaemia will be discussed with timing information and order of magnitude rules for similar investigations. The numbers of causes per person in New Jersey will be summarized in several tables. If time permits, the further analysis of abstracts from this data will be illustrated by three short examples: conventional statistical analysis, a computationally intensive method, and an application of artificial intelligence technique.

The Effects of Heavy Tailed Distributions on the Two-Sided K-Sample Test

Henry D. Crockett
M. M. Whiteside

University of Texas at Arlington

This paper presents the problem that the k-sample Smirnov test has in discriminating the ranking of samples from heavy tailed probability distributions. This is accomplished by performing a multifactored simulation on samples from univariate Cauchy and double exponential distributions. The test results for 1000 tests are presented for each of seven levels of variance, and five scalar offsets for both distributions.

Recent Progress in Algorithms and Architectures for Time Series Analysis

George Cybenko
Department of Computer Science
Tufts University
Medford, MA 02155

617-381-3214

ABSTRACT

This talk will survey research in the 1980's on fast algorithms and computer architectures for time series analysis, especially from the signal processing perspective. A combination of novel algorithms and new technologies are making complex computations not only feasible but performable in real-time by the early 1990's. The talk focuses on techniques involving matrix problems such as eigenvalue, singular value and structured linear system solving. This progress has had added powerful new tools to the time series analyst's collection of techniques.

Chernoff Faces: A PC Implementation

by

Mohammad Dadashzadeh, Ph.D.

Department of Management Science & Information Systems
University of Detroit
4001 W. Mc Nichols
Detroit, MI 48221
(313) 927-1237

ABSTRACT

The Chernoff faces is a well-known method for graphical representation of multivariate data in which every multivariate observation is visualized as a computer-drawn face. As in other techniques for graphical representation of multivariate data, the objective is to assist the investigator in quickly comprehending relevant information in order to apply appropriate statistical analysis. In this paper we present a flexible implementation of Chernoff faces on the IBM PC. The program is written in BASIC and the faces are drawn on the IBM PC's color/graphics screen. Our contribution by this flexible PC implementation of Chernoff faces is to make a rather useful tool more readily accessible to the statisticians for experimentation and possible refinement.

A Numerical Comparison of EM and Quasi-Newton Type Algorithms for Finding MLE's for a Mixture of Normal Distributions

John W. Davenport
Margaret Anne Pierce
Richard J. Hathaway

Georgia Southern College

Calculating maximum likelihood estimates for a mixture of normal distributions is one of the most computationally intensive problems in parametric estimation. Maximizing the corresponding likelihood function is complicated by singularities and numerous spurious maximizers. Currently the most popular technique for finding the particular (local) maximizer of the likelihood function that has good estimation properties is the EM (Expectation Maximization) algorithm. While this iterative algorithm is extremely reliable and usually finds the "good" maximizer from most reasonable initial guesses, it is very slow in cases where the overlap between component normal distributions is great. Another approach, which is faster though thought to be less reliable, is to directly maximize the likelihood function using a (locally) fast iterative algorithm based on some variant of Newton's method. The disadvantage with these quasi-Newton methods is that sometimes the estimate obtained is very dependent on the initial guess used. This paper presents some preliminary numerical results indicating the relative strengths and weaknesses of the EM and quasi-Newton approaches found by testing several methods on a variety of mixture estimation problems. Comparisons made include the computational efficiency and the reliability of the approaches tested. The ultimate goal of this research is to learn how the two basic approaches can be hybridized in order to achieve a method that is both quickly convergent and reliable.

Copy available to DTIC does not permit fully legible reproduction

OPTIMIZING LINEAR FUNCTIONS OF RANDOM VARIABLES HAVING A JOINT MULTINOMIAL OR MULTIVARIATE NORMAL DISTRIBUTION

Abstract

by

JOSEFINA P. DE LOS REYES

A computer method to find vectors $\underline{s}$ that minimize $G(\underline{s}) = \sum_{i=1}^{r} c_i s_i$ ($c_i > 0$ constants) subject to $P\{v_i \le s_i (i=1, \ldots, r)\} \ge 1-\alpha$ $(0<\alpha<1)$ where $v_1, \ldots, v_r$ have a joint multinomial distribution with parameters n, $p_1, \ldots, p_r$ $(p_i > 0,\ p_1 + \ldots + p_r = 1)$ is obtained by solving the corresponding optimization problem through the usual normal approximation. Thus vectors $\underline{x}$ are sought that minimize $F(\underline{x}) = \sum_{i=1}^{r} b_i x_i$ ($b_i > 0$ constants) subject to $P\{X_i \le x_i (i=1,\ldots,r)\} = \Phi_r(x_i, \rho_{ij}) \ge 1-\alpha$ where $X_1, \ldots, X_r$ have a joint (degenerate) multivariate normal distribution with $E(X_i) = 0$, $Var(X_i) = 1$, $Cov(X_i, X_j) = \rho_{ij} = -\{p_i p_j (1-p_i)^{-1}(1-p_j)^{-1}\}^{\frac{1}{2}}$.

The normal probability integral $\Phi_r(x_i, \rho_{ij})$ is evaluated numerically using known computer quadrature codes as (a) one integral over a simplex S, (b) linear combination of integrals over multidimensional right triangles called "plane orthoschemes," or (c) linear combination of integrals over multidimensional rectangular domains.

The optimization of G and F is accomplished using binomial tables and a bisection method for $r = 2$. A known nonlinear program with the numerical quadrature codes for $\Phi_r(x_i, \rho_{ij})$ works well for $r = 3$. For $r \ge 4$, the many evaluations of $\Phi_r(x_i, \rho_{ij})$ required by the optimization routine make the solution difficult and expensive while theoretically simple and feasible. In this regard, the approximation $\Phi_r(x_i, \rho_{ij}) \cong \Phi(x_1) + \ldots + \Phi(x_r) - (r-1)$, where $\Phi(x)$ is the univariate standard normal integral, is shown to be accurate to within $\pm.005$ for values of x_i such that $\Phi_r(x_i, \rho_{ij}) \ge 0.90$ if $x_i = x$, $p_i = \frac{1}{r}$. For $c_1 = \ldots = c_r = 1$, the required probability vectors $\underline{x}$ minimizing F are tabled and related error curves are graphed for $3 \le r \le 30$.

Robustness Study of Some Random Variate Generators

LIH-YUAN DENG

Department of Mathematical Sciences

Memphis State University

Memphis, TN 38152

ABSTRACT

Empirical study using computer-generated random numbers have been widely used where the mathematics of analyzing a statistical procedure become intractable.

There are several generating methods to produce a random sequence with the given distribution. Most of the methods are based on the generation of independent variate from a uniform random distribution. Comparison of the different generating methods usually is done under the criterion of "efficiency". With the wide availability of the mini-, micro- and personal computers, the cost of computing is reducing dramatically. We will adopt a new criterion of "robustness" to compare the performance of different generating schemes.

They are two basic techniques for generating variates from U(0,1): the congruential methods and feedback shift register methods. None of these is known to generate a "true" random sequence. In this paper, using beta random variate generating methods as an example, we will compare the performances of "robustness" of several generators. It is shown that some methods will perform poorly in the sense that it will quite differ from the specified distribution when the uniform generator fails "slightly".

Similar study has been done for comparing different generating methods of normal, gamma ... distributions. The framework of analytical and empirical comparisons will also be discussed.

Compression of Image Data
Using Arithmetic Coding

by

Ahmed Desoky and Thomas Klein
University of Louisville
Louisville, Kentucky 40292

D

Abstract

Arithmetic Coding has been proposed as being more superior in most respects than the Huffman method. This paper examines Arithmetic Coding as a possible compression technique to reduce storage requirements of image data. Arithmetic Coding models are presented along with their performance in specific applications. Quality measures are discussed in terms of a practical image storage and retrieval scheme.

Summary

As image processing projects become more common on personal computers a need arises to reduce image storage requirements. As an example, the University of Louisville Medical School has a lab which produces dozens of images for analysis daily, each image consuming over 1/2 MByte -- enough data to fill a standard PC hard-disk every week. Only recently have coding techniques existed to reduce this burden. These methods include relative encoding, statistical encoding, tree-based encoders and the aforementioned Huffman coder.

Arithmetic Coding represents a message as an interval of real numbers between 0 and 1. The longer the message, the smaller the interval needed to represent it, and thus the more bits needed to specify the interval. An individual symbol of the message reduces the size of the interval by an amount determined by its probability of occurrence, with a more likely symbol reducing the range by less than an unlikely one, and consequently adding fewer bits to the message.

Both the encoder and decoder know (or can generate) the probabilities of occurrences of the various symbols, and also that the initial range is [0,1). With this in mind, the decoder can deduce the final symbol in the message by the range specified, then work backward to reveal the entire message.

In practice, several factors make implementation of this seemingly simple technique less than trivial. Underflow and overflow propensities and overheads caused by message terminators and word-length constraints affect the performance and efficiency of the method. Minimization of these problems requires careful and tedious attention to detail.

The problem of image compression is, in general, very important and lacks unique solutions. Arithmetic Coding, though displaying admirable performance characteristics, appears to be less than an accepted method. A final goal of this paper would then be to examine Arithmetic Coding in detail sufficient to appreciate its effective uses and expose its inherent limitations.

AN L1 ASYMPTOTICALLY OPTIMAL KERNEL ESTIMATE

Luc Devroye
School of Computer Science
McGill University

ABSTRACT

Let f_{nH} be the Parzen-Rosenblatt kernel estimate of a density f on the real line, based upon a sample of n iid random variables drawn from f, and with smoothing factor H depending upon the data. Among other things, we study a fully "automatic" method for picking H such that for a large class of densities, and for any fixed $\varepsilon>0$,

$$\limsup_{n\to\infty} \frac{E(\int |f_{nH}-f|)}{E(\inf_h \int |f_{nh}-f|)} \leq 1+\varepsilon \quad \text{as } n\to\infty ,$$

where f_{nh} is the kernel estimate with smoothing factor h. The H is obtained simply by minimizing $\int |f_{nh}-g_{nh}|$ where g_{nh} is a kernel estimate with a carefully picked kernel depending upon ε and the kernel of f_{nh} only.

Keywords and phrases.

Density estimation. Asymptotic optimality. Nonparametric estimation. Strong convergence. Kernel estimate. Automatic choice of the smoothing factor.

AMS 1980 Subject Classifications.

62G05, 62H99, 62G20.

Author's address: School of Computer Science, McGill University, 805 Sherbrooke Street West, Montreal, Canada H3A 2K6.

Submission for the 20th Symposium on the Interface of Computing Science and Statistics.

Please reply to: Dr. Allen Don
77 Mill Spring Road
Manhasset, New York 11030

A THEORY OF QUADRATURE IN APPLIED PROBABILITY: A Fast Algorithmic Approach

Allen Don, Ph.D

Computer Science Department
Long Island University
Brookville, New York

The integral representation of the moments of a useful class of probability density functions is cast in a canonical form in terms of Gauss-Laguerre quadrature. This transforms the continuous integration into a sum of discrete terms, effectively removing the integral sign and exposing the parameters to numerical investigations. This allows moments from data to be related to the unknown parameters via a system of non-linear equations. This system is easily and quickly solved for the unknown parameters by any of the numerous non-linear equation algorithms available for personal computers and main-frames. In addition, the factorials and gamma functions found in closed form theoretical moment expressions and in density functions are discretized in the same manner, enabling unknown parameters within the arguments of the gamma to be included in numerical searches. A dominant ratios method is introduced for determining initial conditions for the system of non-linear equations to overcome the notable lack of convergence found in non-linear system algorithms when initial conditions are not well-chosen. The notion of finite interval quadrature leads to a correction factor that, with repeated integration-by-parts, becomes an accurate representation of truncated moments with the quadrature terms vanishing. The theory is demonstrated by application to reliability problems, providing a fast algorithmic approach rather than the usual graphical approach to parameter identification of density functions both for truncated and for full data.

Additive Principal Components: A method for estimating equations with small variance from multivariate data

Deborah Donnell
Bell Communications Research

Additive Principal Components are a generalization of linear principal components, where the usual linear function, $a_i X_i$, defining the linear principal component, $\sum_i a_i X_i$, is replaced by a possibly non-linear function, $\phi_i(X_i)$, to form an additive principal component $\sum_i \phi_i(X_i)$. The analogy to the smallest linear principal conponent is investigated. The functions ϕ_i can be estimated by iterative application of a scatterplot smooth. This algorithm is equivalent to a power method of estimating eigenfunctions.

The smallest additive principal components describe nonlinear structure in a high dimensional space. Consequently it is difficult to interpret the estimated functions in terms that are meaningful for the data analyst. For the additive principal component, the task of interpretation is almost intractable without tools for real time graphical interaction. With these tools, a pleasingly direct method for interpretation of the functions in terms of the original variables is possible.

The additive principal component will be defined and the estimation algorithm described. The graphical methodology necessary for interpretation of the results will then be described with the aid of real examples.

MAXIMUM ENTROPY AND THE NEARLY BLACK IMAGE

David Donoho
University of California at Berkeley
and
Iain Johnstone
Stanford University

The maximum entropy estimation principle has been used to derive a non-linear image restoration method intended for use when it is known that the underlying scene is necessarily non-negative. It has been used with success in fields ranging from radio astronomy to spectroscopy. Many of the successful applications have occurred in settings where the scene is positive on a sparse set and is otherwise mostly zero ("black"). This paper begins a quantitative comparison of the maximum entropy method with some other positivity-preserving competitors in some idealised models using a mean-squared error criterion. The simplest situation is that of a "signal plus noise" model. Camparison of estimation methods over a class of "nearly black" images can be cast as a restricted minimax problem. The worst case mean squared error (MSE) for the maximum entropy method, as well as the benchmark minimax MSE must be computed numerically for the fractions of non-black pixels of main interest here. Application of some decision theory significantly reduces the complexity of the necessary computation. It turns out that MEM does indeed make significant gains over the best linear estimator, but that it does not get close to the minimax bound. Indeed, a minimum L1 method, obtained by replacing the entropy functional by the L1 norm performs significantly better numerically. These numerical results are confirmed by an asymptotic analysis that matches the numerics almost exactly at the small non-black fractions at which the computational cost becomes unmanageable. Time permitting, some conjectures concerning the extension of these results to the more complex settings of more general inverse problems will be mentioned.

The Elimination of Quantization Bias Using Dither

Douglas M. Dreher and Martin J. Garbo

HUGHES AIRCRAFT COMPANY

ABSTRACT

This paper presents a method for recovering the decimal precision of a non-observable variable that has been quantized. The technique involves the addition of a random variate (dither) from a uniform distribution to the variable prior to quantization. It then shows the conditions under which the expectation of the dithered quantization function equals the value of the variable in question. An expression for the variance of the dithered quantization function is also derived. The results are then generalized to the multiple-quantization case. Examples involving computer communication are presented which show the application of this technique to reduce the magnitude of bias error caused by roundoff.

GRAPHICAL REPRESENTATIONS OF MAIN EFFECTS AND INTERACTION EFFECTS IN A POLYNOMIAL REGRESSION ON SEVERAL PREDICTORS.

William DuMouchel
BBN Software Products Corporation

The table of coefficients from a polynomial regression analysis having several predictors is hard to interpret because its focus is on the terms in the fitted equation, rather than on the variables used to define those terms. Methods for graphically comparing the effects of each predictor to each other and to the residuals will be introduced and discussed. The techniques are easy to implement and to interpret, and have been generalized to provide graphical summaries of interaction effects.

RECURSIVE METHODS IN TIME-SERIES ANALYSIS

by Quang Phuc Duong

Management Sciences Consulting - Bell Canada, Montreal, Canada

ABSTRACT

Recursive methods have always played an important role in the analysis of Time-Series data, and that for all three main stages of the modeling exercise: identification, estimation and prediction. In addition to the well-known Levinson-Durbin and Kalman Filter algorithms, recent developments, mostly in the field of Control Engineering, have been useful in obtaining efficient estimation methods for the general class of ARMA models through the so-called Innovation approach. This paper reviews the main ideas behind these methods, and then focuses on the problem of estimating the parameters of a Moving Average process; some new concepts are introduced, and it is shown that the resulting algorithm parallels that of the Levison-Durbin algorithm. Other important applications of the algorithm in Time-Series Analysis and other statistical fields are also briefly discussed.

Keywords: Recursive Algorithms Levinson-Durbin algorithm; Innovation Process; spectral Density; Log Autocorrelation.

TESTING MULTIPROCESSING RANDOM NUMBER GENERATORS

Mark J. Durst
Lawrence Livermore National Laboratory

Standard system software on current multiprocessing computers generates pseudo-random numbers which are not reproducible; i.e., different runs will produce different numbers. To preserve reproducibility, multiprocessing random number generators (RNG's) have been proposed. Such generators provide many streams, each of which consists of the numbers to be used by a specific task. These streams should appear individually to be i.i.d. U[0,1], and they should appear to be mutually independent. Suggestions for such generators include deterministically splitting the sequence of a given RNG into substreams, selecting "random" starting points for each substream in a reproducible way, and attempting to create truly distinct streams for each task.

While some theory for such generators can be developed, empirical testing is still important. Standard empirical tests can be used to assure the quality of the individual streams. We discuss some methods for testing whether the streams appear mutually independent. Fixed-dimensional tests which have been used "longitudinally" to test single streams can be used "latitudinally" to test independence of streams. Uniformity tests, permutation tests, and partition tests can be used to test a handful of streams, and collision tests can be used to test about twenty streams. Tests without fixed dimensionality (runs tests, coupon collector's tests, gap tests) can be used latitudinally on a very large number of streams, but a more effective use is to modify the tests slightly to fix their maximum dimensionality. Fourier transforms can be used to derive multiple comparisons tests for cross-correlations and cross-periodogram tests. These are particularly useful in detecting unexpected overlaps of streams. As all these tests involve a great deal of computation, efficient experimental designs for the testing of many streams must be developed.

An Approach for Generation of Two Variable Sets with a Specified Correlation and First ans Second Sample Moments

Mark Eakin
Henry D. Crockett

University of Texas at Arlington

Certain simulations require the generation of correlated variables with prespecified first and second moments. The first step involved the random generation of two standardized variables. Second, the first variable was replaced by a linear combination of the two variables such that the correlation coefficient of the linear combination and the second variable is specified. The variables can then be adjusted to give the required first and second sample moments without modifying the correlation equations.

Asynchronous Iteration

William F. Eddy
Mark J. Schervish
Carnegie Mellon University

An *asynchronous iteration* is an iterative method in which the succesive iterations are not necessarily performed sequentially. Such methods are particularly well-suited to parallel/distributed systems in which several iterations can be performed simultaneously, but not necessarily synchronously. Baudet (1978) and Mitra (1987) prove results concerning the convergence behavior of asynchronous iterative methods for various types of problems. Their results concern the worst case behavior of the method and require conditions on both the behavior of the iterative process and the specific problem being solved. We explore stochastic versions of these results in two specific examples. The examples are

1. Finding the eigenvalues of a large matrix by Gauss-Seidel iterations; and

2. Random affine mappings for producing fractal-like images.

We implement asynchronous iteration on a parallel/distributed system consisting of powerful workstations as described by Eddy and Schervish (1986).

References

Baudet, G.M. (1978). Asynchronous iterative methods for multiprocessors, *Journal of the Association for Computing Machinery*, **25**, 226-244.

Eddy, W.F. and Schervish, M.J. (1986). Discrete-finite inference on a network of VAXes. In *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, T. J. Boardman, Ed. 30-36.

Mitra. D. (1987). Asynchronous relaxations for the numerical solution of differential equations by parallel processors, *SIAM Journal of Scientific and Statistical Computation*, 8, s43-s58.

Determining Properties of Minimal Spanning Trees by Local Sampling

William F. Eddy
and Allen McIntosh
Carnegie-Mellon University
and Bell Communications Research

Let α_{nid} be the fraction of vertices of degree i in a minimal spanning tree on a random sample of n points in d dimensions. Steele, Shepp and Eddy (1987) show that as n increases α_{nid} converges with probability one to an unknown constant α_{id} independent of the sampling distribution. They perform a small scale simulation experiment to determine α_{i2}, $i = 1, \ldots, 5$ by estimating α_{ni2} for increasing values of n when points are distributed uniformly in the unit square. Here, we estimate α_{id} directly by systematically sampling the neighborhood of a particular point of the Poisson process with constant intensity in d dimensions. We discuss a number of techniques used in order to avoid generating large samples $n > 10^5$. We also describe our attempts to estimate $\ldots_{id}$, the number of edges in the minimal spanning tree path between a point and its i^{th} nearest neighbo

ON COMPARATIVE ACCURACY OF MULTIVARIATE NONNORMAL RANDOM NUMBER GENERATORS

Lynne K. Edwards
Department of Educational Psychology
University of Minnesota
Minneapolis, MN 55455-0211

E

Abstract

There are two easily accessible methods of generating multivariate nonnormal distributions using the IMSL. They are: a multivariate extension of Fleishman's (1978) method with an intermediate correlation matrix adjustment and a contamination method. Neither of them can produce all possible combinations of marginal skew and kurtosis, but these methods have an advantage over generating the known extreme distributions when the generation of multivariate nonnormal distributions with specified intercorrelations and specified marginal moments is required to simulate a plausible situation. The MSE for the four marginal moments and for the intercorrelations were compared between these two methods. The Fleishman-type method produced sample correlations much closer to the parameters than the contamination method but the reversed trends were found among the marginal moments.

"Derivative estimation by Polynomial-trigonometric regression"

by

R.L. Eubank, Southern Methodist University
and
Paul Speckman, University of Missouri-Columbia

Abstract

Let μ be a smooth function defined on an interval $[a,b]$ and suppose that $y_1, \ldots, y_n$ are uncorrelated observations with $E(y_j) = \mu(t_j)$ and $Var(y_i) = \sigma^2$, $j=1,\ldots,n$, where the t_j are fixed in $[a,b]$. Estimation of μ and its derivatives by regression on trigonometric and low order polynomial terms is considered. The polynomial terms are shown to adjust for the boundary bias problems known to be suffered by regression on trigonometric terms alone, and the resulting estimate of μ has asymptotics competitive with other nonparametric methods. In addition, if the observations are equally spaced, derivative estimates obtained by this method are competitive with other methods.

ELECTRONIC MAIL - A VALUABLE AUGMENTATION TOOL FOR SCIENTISTS

Elizabeth Feinler
SRI International
Network Information Systems Center
Menlo Park, California 94025
(Electronic Mail: FEINLER@SRI-NIC.ARPA)

ABSTRACT

Most scientists today have access to personal computers, work stations, or mainframe computers in the course of their work. Many of these computers also support electronic mail which can be used to augment the exchange of ideas among researchers. Electronic mail is easy to use and can serve as a transport mechanism for sending data and information quickly and efficiently across networks to other scientists or to other computers. Some of the electronic mail services and programs currently available to scientists are outlined and ways in which they can effectively use electronic mail in their work is discussed.

INFORMATION SYSTEMS AND STATISTICS

Nancy Flournoy
National Science Foundation

The accessibility of high dimensional data presents new challenges to the Statistical Consulting Community. Attention to the orgainization of such data results in a novel environment, rich with opportunities for extending the frontiers of the Decision Sciences. Such a data environment will be described and consequent new statistical methods which are needed will be sketched.

A Comparison of Spearman's Footrule and Rank Correlation Coefficient With Exact Tables and Approximations

LeRoy Franklin
Indiana State University

ABSTRACT:

Given two rankings of n objects a widely used nonparametric measure of association between the rankings is Spearmans-ρ given in unnormalized form as S where

$$S(p,q) = \sum_{i=1}^{n} (p_i - q_i)^2 .$$

However an equally simple but neglected competitor is Spearman's Footrule (1904) and is given in unnormalized form as

$$D(p,q) = \sum_{i=1}^{n} | p_i - q_i | .$$

Diaconis and Graham in a 1977 paper in the *Journal of the Royal Statistical Society* recently renewed interest in D by establishing a limiting normal distribution. Ury and Kleincke in a 1979 paper in Applied Statistics tabulated the exact c.d.f. for D for n=2(1)10 and gave an approximate table for n=11(1)15 generated by Monte Carlo simulation. They also conjectured about the rate of convergence of D and whether an improvement in approximation could be obtained by using a +1 continuity correction factor as is used for Spearman's Rho.

This paper presents exact tables of Spearman's Footrule for n=11(1)18 using computer intensive calculations of the exact permutation distribution. This was done using a specialized program utilizing both permutations and combinations to achieve several orders of magnitudes of increase in CPU processing speed over "direct approach" calculations.

Then for both Spearman's Footrule and Spearman's Correlation Coefficient the maximum differences between the exact c.d.f. and the normal approximation is given as well as the maximum difference between the exact c.d.f. and the normal approximation with correction for continuity. Comparisons are made and graphs of the differences in the c.d.f.'s are provided for representative values of n.

Fitting Functions to Noisy Scattered Data in High Dimensions*

JEROME H. FRIEDMAN

Department of Statistics

and

Stanford Linear Accelerator Center
Stanford University, Stanford, California, 94305

ABSTRACT

Consider an arbitrary domain of interest in n-dimensional Euclidean space and an unknown function of n arguments defined on that domain. The value of the function—perhaps perturbed by additive noise—is given at some set of points. The problem is to find a function that provides a reasonable approximation to the unknown one over the domain of interest. A new approach is presented for the practical solution to this problem. This method, based on adaptive splines, appears to be able to provide smooth, accurate, parsimonious, and interpretable approximations to a wide variety of functions of a multivariate argument.

* Work supported by the Department of Energy, contract DE-AC03-76SF00515.

Dimensionality Constraints on Projection and Section Views of High Dimensional Loci

George W. Furnas
Bell Communications Research

A basic theoretical limitation is shown for the two general graphical techniques for constructing geometric views of high dimensional loci: PROJECTION and SECTION (called "conditioning" in statistical contexts) . Basisically, projections can only easily display aspects of structure that are of low dimensionality. Sections, i.e, intersections of linear subspaces with a locus, can easily display structure of only low CO-dimensionality (and hence high dimensionality). Fortunately, compositions of Section and Projection can display aspects of structure of any intermediate dimensionality. These assertions are proven for fundamental idealization of loci that are arbitrary affine subspaces of a high dimensional space. The issues introduced by finite extent, by curvature, by sampling and by error noise are then discussed, basically in terms of notions of scale. Two examples of using the Projection & Section composition technique are given, examining the structure of high-dimensional objects embedded in a six-dimensional space.

F

BIAS OF ANIMAL POPULATION TREND ESTIMATES

Paul H. Geissler and William A. Link

U.S. Fish and Wildlife Service, Patuxent Wildlife Research Center, Laurel, Maryland

The trend (rate of change) of animal populations is often estimated as

$$\frac{\sum_i A_i \hat{c}_i(\bar{y}+1)}{\sum_i A_i \hat{c}_i(\bar{y})} = \frac{\sum_i A_i \hat{\alpha}_i \hat{\beta}_i}{\sum_i A_i \hat{\alpha}_i}$$

where i indexes sampling units, A is the stratum area, $\hat{c}$ is the predicted count of animals, and $\bar{y}$ is the mean year ($\bar{y}=0$). Counts are estimated using the model

$c_i(y) = \alpha_i \ \beta_i^y \ \theta_{ij} \ \epsilon_{ik}$ where α, β, and θ are the intercept, slope, and observer effect parameters and ϵ is the error. Parameters are estimated by means of linear regression on the logarithmic scale using the unbiased estimation techniques of Bradu & Mundlak.

The bias of this estimator was studied using a factorial simulation experiment with lognormal, Poisson, and negative binomial distributed counts. Bias increases sharply with increasing count variance. Increasing the number of years reduced the bias but increasing the sample size had no discernible affect on the bias. Including observer effects reduces the effective number of years. The direction of the trend had no apparent affect on the bias. The bootstrap was ineffective in reducing the bias. The use of reduced mean square error estimation techniques instead of Bradu & Mundlak's techniques was found to increase the bias.

MIXTURE EXPERIMENTS AND FRACTIONAL FACTORIALS USED TO TAILOR LARGE-SCALE COMPUTER SIMULATIONS

T. K. Gardenier, Ph.D.
TKG Consultants, Ltd.
301 Maple Avenue West, Suite 100
Vienna, VA 22180

Large scale computer simulations are in widespread and growing use in government, business and science. Within the Department of Defense, the use of simulation is particularly crucial because the real-world scenario of the battle cannot be replicated. Environmental and health simulations for risk assessment have complex determinants of pollution and target sites. Large number of parameters may initially appear to be needed in simulations. Experiment designs, and optimization achieved through respense surgace methodology, can reduce the final set of parameters in simulations to an efficient minimum.

The objective of this paper is to present the use of several experiment design procedures, including fractional factorials, mixture experiments with constrained optimization, and Placket-Burman designs based on Hadamard matrices as pre-processors to computer simulations. The methods have been used by the author to (a) minimize the number of computer runs, (b) conduct an input-output analysis of model subroutines and measures of merit, (c) check for computational model validity, (d) design interactive graphical evaluation schemes for the simulation developer and user. These use of these experiment designs as pre-processors resulted in cost-savings as well as efficiency for the types of simulations used in battle-management

Abstract

Acceleration Methods for Monte Carlo Integration in Bayesian Inference

John Geweke, Duke University

Methods for the acceleration of Monte Carlo integration with n replications in a sample of size T are investigated. A general procedure for combining antithetic variation and grid methods with Monte Carlo methods is proposed, and it is shown that the numerical accuracy of these hybrid methods can be evaluated routinely. The derivation indicates the characteristics of applications in which acceleration is likely to be most beneficial. This is confirmed in a series of examples, in which these acceleration methods reduce n and the computation time required to achieve a given degree of numerical accuracy by up to several orders of magnitude. The methods are especially well suited to vector processors, and on such processors substantial further decreases in computation time are achieved. It is shown that without acceleration the standard deviation of the numerical error in Monte Carlo integration is $O(1/nT)$, and if antithetic acceleration is incorporated it is $O(1/nT^2)$. It is conjectured that with the incorporation of grid methods this standard error is $O(1/n^2T)$, and that with both antithetic variation and grid methods it is $O(1/n^2T^2)$.

G

Higher Order Functions in Numerical Programming

David S. Gladstein

ICAD, Inc.
1000 Massachusetts Avenue
Cambridge, Massachusetts 02138

Conventional algebraic programming languages like C, Pascal, and Fortran have statically defined functions and procedures, which are completely established by the time a program is compiled and linked. In contrast, Lisp, Scheme, and other symbolic programming languages consider functions to be *first class objects*, meaning that they can be used as data and created at run time. Functions which map functions to functions are said to be of *higher order*.

Higher order functions arise naturally in many ways. As a case study, I consider the sequential analysis problem of computing certain confidence intervals using a probability $A(x, i)$:

$$A(x,i) = \begin{cases} 1 - \Phi(x-\theta), & \text{for } i = 1; \\ \int_a^b f_{i-1}(y)\big(1 - \Phi(x-y-\theta)\big)\,dy & \text{for } i \geq 2, \end{cases}$$

where

$$f_1(x) = \begin{cases} \phi(x-\theta), & \text{for } a \leq x \leq b; \\ 0, & \text{otherwise,} \end{cases}$$

and

$$f_i(x) = \int_a^b f_{i-1}(y)\phi(x-y-\theta)\,dy \qquad \text{for } i \geq 2.$$

(ϕ is the standard normal density, Φ is the standard normal distribution, and a, b, and θ are fixed parameters.)

A naïve implementation runs in time exponential in i, because each evaluation of f_i requires integrating a function involving f_{i-1} over the interval $[a, b]$, and so on until f_1. To achieve run time linear in i, we must introduce the complication of saving (or *cacheing*) each value of each function f_i as it is computed.

Implementing this calculation in C is very tedious, and results in much code tailored to the specific problem. I show how Lisp's ability to generate functions at run time results in a program with several desirable properties:

1. The structure of the program mirrors the mathematical formulation of the problem. The use of cacheing functions increases the size of the routine which calculates $A(x, i)$ by only one function call.
2. All integration is performed by a single, general purpose integration routine. This routine is used to map a function f to another function $F(a, b) \equiv \int_a^b f(t)\,dt$.
3. All cacheing functions are produced by a general purpose function, which maps a function f onto a cacheing version which produces the same results but caches all computations. The cacheing version is as easy to deal with as the original function.
4. The array of (cacheing) functions $\{f_1, f_2, \ldots f_i\}$ is simply computed from their definition. i can be arbitrarily large, subject only to memory constraints.

The complete Common Lisp source program for the sequential confidence interval problem is presented, with a discussion of how the implementation differs in C.

Performance comparisons between a Common Lisp version running on a Lisp machine and a C version running on various configurations of personal computers are presented.

MOVING WINDOW DETECTION FOR 0-1 MARKOV TRIALS

Joseph Glaz*, Philip C. Hormel** and Bruce McK. Johnson***
University of Connecticut and CIBA-GEIGY Corporation

ABSTRACT

Let $X_1, X_2, \ldots$ be a sequence of 0-1 Markov trials. The random variable X_i represents the number of signals that were detected at the end of the ith discrete-time interval. The k-out-of-m moving window detector generates a pulse whenever k or more signals are detected within m consecutive discrete-time intervals. Define $M_{k,m}$ to be the waiting time for detection using a k-out-of-m moving window detector. In this article we derive Bonferroni-type and product-type approximations for the distribution of $M_{k,m}$, which in turn yield approximations for $E(M_{k,m})$ and $VAR(M_{k,m})$. These quantities play an important role in the design and analysis of the k-out-of-m moving window detection procedure. Applications to the theory of radar detection and quality control (zone tests) are discussed.

* Joseph Glaz is Associate Professor, Department of Statistics, University of Connecticut, Storrs CT 06268.

** Philip C. Hormel is Biostatistician, Marketing Clinical Support, CIBA-GEIGY Corporation, Summit NJ 07901.

*** Professor Bruce McK. Johnson was with the Department of Statistics, University of Connecticut, Storrs CT 06268. Regrettably, he passed away on November 4, 1986.

Abstract

Software for Bayesian Analysis : Current Status and Additional Needs

Prem K. Goel
The Ohio State University
Columbus, Ohio 43210, U.S.A.

We make an attempt to provide comprehensive information about the existing software for data analysis within the Bayesian paradigm. The paucity of programs seems to indicate that the Bayesian software available for widespread use is still in its infancy. We have a long way to go before a general purpose Bayesian Statistical Analysis Package is made available. Alternatives for reaching this goal quickly are presented in the concluding section.

Paper

Computational Requirements of Inference Methods in Expert Systems: A Comparative Study

by

Ambrose Goicoechea
School of Information Technology and Engineering
George Mason University

G

Abstract

This paper presents a detailed comparative study of six major, leading methods for inexact reasoning: (1) Bayes' Rule, (2) Dempster-Shafer theory, (3) Fuzzy Set Theory, (4) MYCIN Model, (5) Cohen's System of inductive probabilities, and (6) a class of non-monotonic reasoning methods. Each method is presented and discussed in terms of theoretical content, a detailed numerical example, and a list of strengths and limitations. Purposely, the same numerical example is addressed by each method to be able to highlight the assumptions, knowledge representation and computational requirements that are specific to each method. Guidelines are offered to assist in the selection of the method that is most appropriate for a particular problem.

KEY WORDS: Inference models, expert systems, imperfect knowledge, uncertainty, decision support systems, inference network, evidential reasoning.

Presented at the Twentieth Symposium on the Interface of Computing Science and Statistics, Reston, Virginia, April 21-23, 1988.

A Simulated Annealing Approach to Mapping DNA

Larry Goldstein and Michael Waterman
University of Southern California

The double digest mapping problem that arises in molecular biology is an NP complete problem that shares similarity with both the travelling salesman problem and the partition problem. Sequences of DNA are cut at short specific patterns by one of two restriction enzymes singly and then by both in combination. From the set of resulting lengths, one is required to construct a map showing the location of cleavage sites. In order to implement the simulated annealing algorithm, one must define appropriate neighborhoods on the configuration space, in this case a pair of permutations, and an energy function to minimize that attains its global minimum value at the true solution. We study the performance of the simulated annealing algorithm for the double digest problem with a particular energy function and a neighborhood structure based on a deterministic procedure for the travelling salesman problem.

SPACE BALLS!
OR
ESTIMATING DIAMETER DISTRIBUTIONS OF POLYSTYRENE MICROSPHERES

CHARLES HAGWOOD AND SUSANNAH SCHILLER
NATIONAL BUREAU OF STANDARDS
GAITHERSBURG, MD 20899

POLYSTYRENE MICROSPHERES, WITH NOMINAL DIAMETERS IN THE RANGE OF 1 TO 30 MICRONS, WERE MANUFACTURED IN SPACE ON THE SHUTTLE CHALLENGER AND ARE CERTIFIED BY THE NATIONAL BUREAU OF STANDARDS AS STANDARD REFERENCE MATERIALS; THEY PROVIDE AN IMPORTANT TOOL FOR CALIBRATING INSTRUMENTS THAT ARE USED TO EXAMINE VERY SMALL PARTICLES. IN ORDER TO BE USEFUL, THEIR DIAMETER DISTRIBUTIONS MUST BE WELL-CHARACTERIZED. ONE MEASUREMENT TECHNIQUE PROPOSED IS TO FORM CLOSELY PACKED HEXAGONAL ARRAYS ON A MICROSCOPE SLIDE WITH THE SPHERES, MEASURE THE ROW LENGTHS, AND IMPUTE THE DIAMETERS FROM THESE. THE OBVIOUS DIAMETER ESTIMATE IS THE ROW LENGTH DIVIDED BY THE NUMBER OF SPHERES IN THE ROW. HOWEVER, BECAUSE THE DIAMETERS ARE NOT IDENTICAL, THERE ARE ALWAYS AIR GAPS IN THESE ARRAYS WHICH INFLATE THE DIAMETER ESTIMATES. THESE AIR GAPS CANNOT BE MEASURED BY THE MICROSCOPE, NOR CAN THEY BE MODELLED MATHEMATICALLY. THEREFORE, OUR APPROACH TO THIS ESTIMATION PROBLEM IS TO SIMULATE ARRAYS OF THE SPHERES AND DETERMINE THE BEHAVIOUR OF THE AIR GAPS. METHODS OF SEQUENTIAL ANALYSIS ARE USED TO DERIVE ESTIMATES OF THE MEAN DIAMETER AND ITS VARIANCE.

TIME SERIES IN A MICROCOMPUTER ENVIRONMENT

John Henstridge, Numerical Algorithms Group

Microcomputers provide a major challenge to statistical software writers not only because of their small memory and relatively poor compilers compared with mainframes but also because users have come to expect a very high standard of "user friendliness". This standard has been set by business oriented software such as wordprocessors and spreadsheets and compared with these most mainframe statistical software stands up very poorly. Partly this problem stems from the tradition in statistical computing for packages to be highly portable and hence make no use of special facilities in any single computer.

This challenge was encountered when transfering a major times series package TSA onto IBM type personal computers.[1] As well as the obvious need to give the package a more screen oriented appearence it was found desirable to develop an environment especially for the most difficult time series problem - time domain and tranfer function model selection and fitting. This entailed the package keeping records of the history of the fitting process and enabling the user to recall details of statistical importance so that models could be readily compared and assessed. The numerically intense nature of most times domain model fitting and the relative slow speed of personal computers also demanded that the package make efficient use of any information previously gained about the series being modeled and previously fitted models.

A second area where major enhancement was considered necessary was that of graphics. In particular a blend of default graphical styles for the first time user had to be developed in parallel with a system which gives complete control to the advanced user.

The final result is a highly interactive system which can perform most time series operations in both frequency and time domains in a manner which emphasises the productivity of the statistician using it.

[1] Henstridge, J.D., 1982, *TSA, An interactive package for times series analysis*, NAG, Oxford.

ABSTRACT

A DATA ANALYSIS AND BAYESIAN FRAMEWORK FOR ERRORS-IN-VARIABLES

John H. Herbert, Department of Energy

More than fifty years ago Ragnar Frisch, the first Nobel prize winner in economics, set forth graphical and statistical procedures for determining the effect of errors-in-variables on estimated coefficients in a regression analysis. The procedures were recommended on heuristic grounds but their statistical properties were not delineated. The procedures were also viewed as computationally prohibitive.

Patefield in a 1981 article in the Journal of the Royal Statistical Society demonstrated that the statistical procedure set forth by Frisch yields maximum likelihood bounds for a true coefficient. Klepper and Leamer in a 1984 Econometrica article extended the procedure within a Bayesian Framework. Stewart in a 1981 article in Statistical Science recommended the collinearity indices that are byproducts of the Frisch errors-in-variables regression procedure as ideal collinearity indices.

In this paper we will first summarize the statistical properties of the Frisch procedure. Then, a relatively simple computational procedure for obtaining solutions will be examined in detail. This computational procedure yields the collinearity indices. Finally, the methodology will be applied to an actual problem with real data to demonstrate the usefulness of the procedure as a tool for a regression analysis.

ABSTRACT

COMPARING SAMPLE REUSE METHODS AT FHA-- AN EMPIRICAL APPROACH

Thomas Herzog
U. S. Department of Housing and Urban Development

The Federal Housing Administration (FHA) recently completed a study of its single-family home mortgage insurance program for investor (i.e., non-occupant) loans. A probability sample of over 6,000 loans was drawn and the results were analyzed using both Bayesian and sample reuse procedures. In this work, we compare the results of the sample reuse methods to each other as well as to the Bayesian method. Finally, Monte Carlo methods are used to simulate the results to see to what extent the same relationships hold under various schemes for generating pseudorandom numbers.

H

ABSTRACT FOR THE 20TH SYMPOSIUM ON THE INTERFACE:
COMPUTING SCIENCE AND STATISTICS 1988, RESTON, VA

INSIDE A STATISTICAL EXPERT SYSTEM:
Implementation of the ESTES expert system

Paula Hietala
University of Tampere
Department of Mathematical Sciences/Statistics
P.O. Box 607, SF-33101 Tampere, FINLAND

Keywords: Expert systems; Rules, Explanation capabilities

Statistical expert systems are an interesting and novel area of statistical computing today (see e.g. Chambers (1981), Gale (1986) and Hietala (1987)). However, the implementations of these systems are often described very cursorily and the reader is left unaware or in doubt of the methods employed as well as of the inner structure of the system. In this paper we consider the implementation of a statistical expert system called ESTES (Expert System for TimE Series analysis) in more detail. The ESTES system is intended to provide guidance for an inexperienced time series analyst in the preliminary analysis of time series, i.e. in detecting and handling of seasonality, trend, outliers, level shifts and other essential properties of time series (Hietala (1986)). Our system is organized so that as much as possible of *knowledge or experience* of the user (about the specific time series being considered) *is exploited*. Even in the case of an inexperienced user he/she may have plenty of useful knowledge concerning the environment of the problem in question. However, if there exists a conflict between the initial results computed by the system and the knowledge elicited from the user, then the ESTES system sets out to carry out more extensive analysis and apply more sophisticated statistical methods. With this kind of organization we strive for minimizing the number of unnecessary reasoning and calculation steps.

The ESTES system has been implemented on Apple Macintosh™ personal microcomputers using Prolog and Modula-2 languages. We have selected *if-then rules* for representing knowledge on properties of time series and their handling. Rules have many desirable features (modularity, incrementability and modifiability, see Bratko (1986)). Rules in our system are either of form: RuleName: *if* condition A *then* conclusion B, or of form: RuleName: *if* condition A *then* action C. The condition part of a rule may be combined (it can contain *and* and *or* operators); moreover, a condition and an action usually include an invisible call to Modula-2 procedures. This kind of rules are easily expressed in Prolog: in fact, they are legal Prolog clauses if we define appropriate operators (e.g. :, if, then). The rule-base of the ESTES system has been organized hierarchically according to (1) the property being considered, (2) the level of analysis process (whether we performing initial or more extensive analysis) and (3) the goal of the analyzing (detecting or handling of the property).

One of the most essential features of an expert system is its ability to explain its own actions. With this in mind, we have paid special emphasis to *the explanation capabilities* of the ESTES system. We do not use Prolog's own trace facility but have built an interpreter on top of Prolog. This interpreter manages the reasoning process of the ESTES system: it accepts questions and finds answers. For example, user can ask 'why' and 'how' questions ("Why the system inquires this fact?", "How the system has reached this conclusion?", see e.g. Bratko (1986)); our system's reply consists of displaying a user-friendly form of its inner inference chain with explanations and justifications of those methods that are used inside the chain.

In the full paper we will describe in detail the *formalisms employed in representing knowledge* and the *structure of our inference engine*. We will also characterize the *interface* between the rule-base part (Prolog clauses) and the statistical part (Modula-2 procedures) of the ESTES system.

REFERENCES:

Bratko, I. (1986). *Prolog programming for artificial intelligence.* Addison-Wesley, Wokingham, England.

Chambers, J.M. (1981). Some thoughts on expert software. *Computer Science and Statistics: Proceedings of the 13th Symposium on Interface.* Springer-Verlag, New York, NY, 36-40.

Gale, W.A. (ed.), (1986). *Artificial Intelligence & Statistics.* Addison-Wesley, Reading, MA.

Hietala, P. (1986). How to assist an inexperienced user in the preliminary analysis of time series: First version of the ESTES expert system. *Proceedings in Computational Statistics (COMPSTAT) 1986, 7th Symposium held at Rome 1986*, Physica Verlag, Heidelberg, 295-300.

Hietala, P. (1987). Statistical expert systems for time series analysis. Paper presented at *The First Conference on Statistical Computing (ICOSCO-I)*, Cesme, Turkey, 30 March - 2 April, 1987.

The Data Viewer: A Program for Graphical Analysis

Catherine Hurley

University of Waterloo

The presentation will contain descriptions of some graphical methods for analyzing multivariate data and their inplementation in the data viewer program. The program produces plots moving in real-time by projecting onto a sequence of user-controlled planes. Multiple plots may be simultaneously controlled, allowing dynamic comparisons of data sets.

The data viewer constructs sequences of planes by interpolating between user-chosen target planes. Following the proposal of Buja and Asimov (1985), the program interpolates along geodesic paths. Available chioces include planes yielding bivariate scatterplots, principal components, or cannonical variable plots.

H

When plots are *linked*, they may be simultaneously controlled and manipulated. With the data viewer's object oriented design, such linked plots are easily constructed. As a consequence, data sets may be compared and related in very general ways.

SPLINE ESTIMATION OF DEATH DENSITY USING CENSUS AND VITAL STATISTICS DATA

John J. Hsieh, University of Toronto

This paper develops a precise method for constructing period life tables through estimation of death density functions using spline method. The paper derives a set of formulas for computing the survival function from the observed cross-sectional death and population data in five-year age groupings. A complete cubic spline is then fitted through the computed survival curve defined on a mesh with n age points as knots. The two endslopes as boundary conditions are determined from observed population and death data using the properties of the lifetime distribution. Death density function is obtained by spline differenciation of the survival function. Hazard function is then obtained as the ratio of the density and survival functions. The article also contains spline integration for computing the person-years lived and the life expectancy as well as interpolation for making a complete life table from the abridged life table so constructed. The complete cubic cardinal spline representation allows best approximation (minimum morm, rapid convergence, etc.) to be simply and stably computed using existing algorithm. The parameters are determined by solving the n+2 systems of n-2 linear equations together with the two boundary conditions for the cardinal spline. The tridiagonal form of the coefficient matrices allows the linear systems to be easily solved using a computer by Gaussian elimination which simplifies to the "Thomas algorithm". Furthermore, the diagonal dominance and symetric characteristic of the matrices guarantee stable results with minimum accumulation of rounding error.

Simultaneous Confidence Intervals in the General Linear Model by Jason C. Hsu

In the general linear model (GLM) $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\underline{Y}$ is a vector of observations, **X** is a known design matrix, $\underline{\beta} = (\beta_1, \dots, \beta_p)$ are unknown parameters, and $\underline{\varepsilon}$ is a vector of *iid* normal errors. Suppose $\underline{\beta}^* = (\beta_1, \dots, \beta_k)$ are of interest ($k \le p$); $\beta_1, \dots, \beta_k$ may be the coefficients in a response surface model, or treatment contrasts in an ANOVA or ANCOVA setting. Consider simultaneous confidence intervals $\{\beta_i \in b_i \pm cs(b_i) \text{ for } i = 1, \cdots, k\}$ where $\underline{b}$ is the least square estimator of $\underline{\beta}^*$ and $s(b_i)$ is the estimated standard deviation of b_i. The exact coverage probability CovProb = $P\{|b_i - \beta_i|/s(b_i) \le c \text{ for } i = 1, \cdots, k\}$, and thus the critical value c, is computable in real time by quadrature if the correlation matrix **R** of $\underline{b}$ satisfies

$$\mathbf{R} = \begin{pmatrix} 1-\lambda_1^2 & & 0 \\ & \ddots & \\ 0 & & 1-\lambda_k^2 \end{pmatrix} \pm \begin{pmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_k \end{pmatrix} (\lambda_1 \cdots \lambda_k) \tag{1}$$

for some $\underline{\lambda} = (\lambda_1, \dots, \lambda_k)'$. In real life **R** rarely satisfies (1), due to covariates and/or missing values. Instead of using Scheffé's projection method or Sidak's inequality to bound CovProb below by $1-\alpha$, we approximate CovProb by replacing the given **R** with the "closest" correlation matrix **R'** satisfying (1). In the case of the + sign, this is equivalent to finding an auxiliary variable b_0 so that $(b_1, \cdots, b_k)$ conditional on b_0 are almost independent, and *conditionally* pretending them to be independent in analogy with Sidak's method. The key is that **R'** is the 1-factor decomposition of the *deterministic* matrix **R**, which can be computed using existing Factor Analysis algorithms for various norms. The case of the – sign, which involves complex integration, can also be handled.

Simulation shows the approximation to be excellent. For comparing treatments when there are covariates (ANCOVA), using a real data set in *Scheffé* for example, variance-reduced simulation estimates of true non-coverage probability α are

Nominal α	Unbiased Estimate of True α	95% Confidence Interval for True α
0.10	0.10 - 0.000025	(0.0991, 0.1008)
0.05	0.05 + 0.000175	(0.0496, 0.0508)
0.01	0.01 - 0.000125	(0.0096, 0.0101)

Improvement over traditional methods is substantial. For a real data set in *Draper and Smith*, for example, the critical value c that determines the half-widths of the confidence intervals are as follows for various methods:

Bonferroni	Sidak	Scheffé	**Proposed**
3.206	3.194	3.919	**2.525**

The MEANS option in PROC GLM of SAS ignores the nuisance parameters $\beta_{k+1}, \dots, \beta_p$ in the user-specified model, in order to guarantee that **R** satisfies (1) in an ANOVA or ANCOVA setting. But the resulting $\underline{b}$ does not estimate $\underline{\beta}^*$ in the user's model, rendering the confidence intervals produced meaningless. This little known error in SAS casts doubts on some published findings (e.g. *Science* 1987, pp. 1110-1113).

The Simulation of Life Tests with Random Censoring

Joseph C. Hudson
GMI Engineering & Management Institute

Abstract

n items are placed on test. Each item ramains on test until either failure or removal from test by a random censoring mechanism independent from the failure mechanism. Such censoring can result from failure of the test apparatus or from failure due to a failure mechanism independent from the one under study. This paper considers the simulation of such a life test under the constraint that the number of items censored is a Binomial random variable with parameters n and p_c, where p_c is the probability of censoring. This allows simulations to be run specifying the expected percentage of censored items.

Simulations are carried out using Weibull, Uniform, Truncated Normal and Truncated Cauchy failure distributions. The censoring distribution is taken to be Exponential. With user-specified failure distribution and probability of censoring, the mean of the censoring distribution is determined so as to enforce the constraint that $P(T_{ci} < T_{fi}) = p_c$. Where T_{ci} and T_{fi} are the censoring and failure times of the i^{th} item, respectively. A failure time and a censoring time are independently generated for each item, with the smaller of these times taken as the time of removal from test.

Details of the implementation are discussed and a validation study is presented. An appendix gives mathematical derivations. The simulation is implemented in Pascal.

VISUAL MULTI-DIMENSIONAL GEOMETRY With APPLICATIONS

Alfred Inselberg * # & Bernard Dimsdale *

* *IBM Scientific Center* & # *Department of Computer Science*

11601 Wilshire Boulevard
Los Angeles, CA 90025-1738

University of California
Los Angeles, CA 90024

A non-projective mapping $R^N \to R^2$ for any positive integer N is obtained from new system of *Parallel Coordinates*. Relations in N variables are portrayed as planar "graphs" having certain properties analogous to the corresponding Hypersurface in R^N. In the plane a *point* $\leftarrow\rightarrow$ *line* duality leads to efficient algorithms for Convex Merge and Intersection of Convex Sets. A line in R^N is represented by $N-1$ planar points and a hyperplane by $N-1$ vertical lines. These enable some geometrical constructions and the representation of polyhedra in R^N. The representation of a class of more general convex and nonconvex hypersurfaces is known. There is an algorithm for constructing and displaying any point interior, exterior or on a hypersurface belonging to these class. Computer Graphics implementations will be shown of:

- the representations,
- algorithms,
- application to *Exploratory Data Analysis in Statistics*, and
- a new *Air Traffic Control System* (i.e. R^4) where the time and space trajectory information is displayed and used in collision avoidance (proximity) and routing.

Knowledge-based Project Management: Work Effort Estimation

V. Kanabar
Department of Mathematics
University of Winnipeg

Knowledge-based techniques are applied to project management work effort estimation and resource selection. The estimating process is one of the most critical and difficult activities in project management. By integrating knowledge-based technology with project management we provide a certain deductive capability that is useful in wortk effort estimation. This paper describes such a model and the statistical techniques used to produce estimates of work effort involved in a project.

MAXIMUM ENTROPY AND ITS APPLICATION TO LINGUISTIC DIVERSITY

By

R. K. Jain

Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, Newfoundland, Canada A1C 5S7

ABSTRACT

The linguistic diversity in large cities are observed to fluctuate in time due to the rapidly changing pattern of immigration. The large changes in ethical population requires proper planning and policies which in turn is dependent upon their prediction of the future representation in the community. In this paper, the principle of maximum entropy is applied to study the linguistic diversity under uncertainty. Numerical example is used to illustrate an algorithm for predicting probability distribution based on the principle of maximum entropy.

Key Words: Maximum Entropy Principle and Linguistic Diversity

Discrete Structures and Reliability Computations

James P. Jarvis
Department of Mathematical Sciences
Clemson University

Douglas R. Shier
Department of Mathematics
College of William and Mary

The computation of the reliability of a system, in terms of the reliabilities of its components, has become increasingly important in assessing the performance of various computer, telecommunication, and distribution networks. For example, in a typical scenario, the edges of a network are assumed to fail randomly and independently with known probabilities and it is required to calculate the probability that the system functions (e.g., supports point-to-point message transmission).

Unfortunately, the computation of most probabilistic measures for general networks is mathematically intractable (i.e., NP-hard). Thus it is fairly unlikely that good algorithms (with time complexity polynomially bounded in the size of the network) can ever be devised. However, it has recently been found that "pseudopolynomial" algorithms are possible for certain network reliability problems: namely, algorithms whose complexity is polynomial in the number of paths or cutsets in the network.

This talk will discuss the role of discrete computation in calculating the "two-terminal" reliability of planar networks (still an NP-hard problem). Specifically, we first discuss data structures for representing, manipulating, and traversing planar graphs. Such structures are then used to develop highly efficient methods for generating paths and cutsets in planar graphs. Finally, certain algebraic structures (lattices) are employed to aid in combining such combinatorial objects (paths, cutsets) to produce the reliability polynomial for planar systems. These methods are applied to some fairly challenging examples from the literature, and representative computational results are presented.

J

AUTOMATIC DETECTION OF THE OPTIC NERVE IN COLOR IMAGES OF THE RETINA

Norman Katz, Subhasis Chaudhuri*, Michael Goldbaum and Mark Nelson**

Dept. of Ophthalmology, *Dept. of Electrical Engineering,
University of California, San Diego
La Jolla, CA 92093

** Radford Company
1755 Homet Road
Pasadena, CA 91106

Detection and identification of objects in retinal images plays an important role in assisting physicians in diagnosing diseases of the eye. Normal objects typically found in these images include blood vessels, the optic nerve and the fovea. Abnormal objects include hemmorhages and lesions. Some progress has already been reported by different researchers in detecting blood vessels in these images. However, little work has been discussed in which the optic nerve is automatically identified. We have developed a method that combines image processing algorithms with Bayesian classification rules to determine the location of the optic nerve in retinal images.

The optic nerve, also known as the optic disk, may be characterized as a bright, elliptically shaped object in the retinal image. However, the detection of the disk is often complicated by the presence of arbitrarily shaped abnormal objects known as lesions. The size, shape, brightness and color of these lesions vary widely among different images, according to the nature and progression of the patient's disease. For this reason, no single characteristic feature can be used to correctly identify the optic disk.

The proposed method includes five classification rules, based on certain physiological properties of the optic disk: (a) size, in terms of major and minor axes of the ellipse; (b) brightness; (c) color; (d) density of edges including both the rim of the disk and the blood vessels within the disk area; (e) presence of large caliber vertically-oriented blood vessels directly above and below the disk. By suitable choice of weighting coefficients, these rules can be combined to determine the maximum likelihood estimate for classification of the disk. This technique has been found to be effective in a large number of retinal images. It is also being incorporated into a system for automatic diagnosis of retinal diseases.

The Use of General Modified Exponential
Curves in Software Reliability Modeling

by

T. M. Khoshgoftaar

Department of Computer Science
Florida Atlantic University
Boca Raton, Florida 33431
Telephone (305) 393-3994

In this paper, we develop a nonhomogeneous Poisson process with a mean value function which has a General Modified Exponential growth curve for the number of detected software errors. This model produces an exponential growth curve (Goel and Okumoto model) and the Logistic and Gompertz models as special cases. It should be recognized that by fitting the Goel and Okumoto model, the Logistic model, or the Gompertz model to a data set of software failures, a prior restriction is being imposed upon the more generalized model. Such restrictions may be inappropriate in any particular application. By fitting the General Modified Exponential model, the power law parameter, p, is estimated by the data and is not constrained, possibly incorrectly, to -1 (Goel and Okumoto model), +1 (Logistic model), or zero (Gompertz model). Therefore, a much wider range of growth curves become available, offering the possibility of finding a more appropriate functional form in any situation.

The parameters of this model are estimated using the maximum likelihood method. Comparisons with other software reliability models are made.

A set of failure data, which was collected from a real time command and control system, is used to fit each model.

ASSESSMENT OF PREDICTION PROCEDURES IN MULTIPLE REGRESSION ANALYSIS

Victor Kipnis

As opposed to the traditional inference a major goal of modern regression analysis is model building, i.e., obtaining a regression equation satisfying some specified criterion. When the purpose of regression analysis is prediction of new observations, model building is usually reduced to selection of a predictor among the class of potential predictors. The paper examines the problem of estimating of the mean squared error of prediction (MSEP) for a linear regression predictor chosen by a given selection procedure. The theory behind the conventional MSEP estimators is not valid when predictor selection and estimation are from the same data. The very selection process affects the distribution of those estimators and, in particular, leads to their substantial bias when the selection effect is not allowed for. To be able to get an adequate estimator we bring in the "procedural approach" and suggest that assessment of the efficiency of a predictor should rest on the assessment of the selection procedure by which this predictor has been chosen, rather than the evaluation of any particular predictor equation. As exact distributional results are virtually impossible to obtain, even for the simplest of common selection procedures, the suggested approach is based on generating bootstrap pseudosamples and applying to them the same selection procedure that was used for the original data. Simulation results comparing MSEP estimators provided by this method with the conventional ones are described. It is also shown that the presented method may help in finding a good predictor.

Numerical Approach to Non-Gaussian Smoothing and Its Applications

Genshiro Kitagawa
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo JAPAN

Recursive formula for filtering and smoothing of general non-Gaussian state space model can be obtained. The formula can be realized by various numerical approximation methods. Thus the analog of the Kalman filter and fixed interval smoothing algorithm can be applied to various time series problems. Some applications of the non-Gaussian state space modeling is also shown.

K

Dynamically Updating Relevance Judgements in Probabilistic Information Systems via User's Feedback

Peter J. Lenk
Barry D. Floyd
New York University

A decision maker's performance relies on the availability of relevant information. In many environments, the relation between the decision maker's informational needs and the information base are complex and uncertain. A fundamental concept of information systems, such as decision support and document retrieval, is the probability that the retrieved information is useful to the decision maker's query. This paper present a sequential, Bayesian, probabilistic indexing model that explicitly combines expert opinion with data about the system's performance. The expert opinion is encoded into probability statements. These statements are modified by the user's feedback about the relevance of the retrieved information to their queries. The predictive probability that a datum in the information base is applicable to the current query is a logistic function of expert opinion and the feedback. This feedback enters the computation through a measure of association between the current query-datum pair with previous, relevant query-datum pairs. When this measure is based on the proportional matching of multiple attributes, the predictive probabilities have a recursive formula that makes the model computationally feasible for large information bases.

Keywords: Decision theory, Bayesian inference, decision support systems, expert systems, document retrieval, probabilistic indexing.

Author: J. Knaub

Organization: Energy Information Administration, Office of Oil & Gas

Title: A Sensitivity Analysis of the Herfindahl-Hirschman Index (HHI)

Abstract:

When comparing the HHI value for a given situation in one time period, to another time period, there is a question as to when one can say a substantial change has taken place. If a small change in a frame often results in a large change in the HHI, then a small change in the HHI may not mean very much. Conversely, if a large change in a frame often results in a small change in the HHI, then one could say a small change in HHI may be very important. (Note that if both of these situations are true, this would be analogous to an hypothesis test where both the Type I and Type II error probabilities are large.) Further, there is the inherent question as to what is a large change and what is a small change. In this paper an attempt is made to answer these questions for given sets of data from the petroleum industry, used by the Energy Information Administration.

Specifically, data were examined for companies by State for a given product. Companies were drawn at random with replacement from the original list of companies for the given State and product. When the same number of companies were drawn as orginally found, the HHI was calculated for this new set of companies. This case, called "unrestricted," is only of passing interest, as a case where the total volume for the State and product must be within, say, five percent of the original total volume is more relevant to this study. Coefficients of variation (CVs) were found (for different numbers of replications). Thus, one could see what changes in the HHI could be expected when companies of the same type, number, and approximately the same total volume are used for each State/product.

AN INTRODUCTION TO CART^{TM}: CLASSIFICATION AND REGRESSION TREES

Gerard T. LaVarnway
Department of Mathematics
Norwich University
Northfield, Vermont

ABSTRACT

The general classification problem may be described as follows: Given a multivariate observation z which is known to belong to (emanate from) one of n possible populations (platforms), determine which population is most likely. The analyst who is performing this classification has a historic data base of observations, for each of which the actual population is known, and has suspicions - in the form of prior probabilities - regarding the likely population of z.

Traditional methods of dealing with this problem often lack flexibility. Observations, for example, are often assumed to be normally distributed. Traditional methods typically cannot deal with observations that contain categorical variables or missing data in a natural way.

The flexible nonparametric approach described in CART (Classification and Regression Trees (1984) Breiman, et al., Wadsworth) will be discussed. The classification rules appear in the form of binary decision trees which are easy to use, understand and interpret.

AN EMPIRICAL BAYES DECISION RULE OF TWO-CLASS PATTERN RECOGNITTION FOR ONE-DIMENSIONAL PARAMETRIC DISTRIBUTIONS

By

Tze Fen Li and Dinesh S. Bhoj

ABSTRACT

In the pattern classification problems, it is known that the Bayes decision rule, which separates two classes, gives a minimum probability of misclassification. In this paper, we assume that the conditional density belongs to any parametric family with unknown parameters and that the prior probability of each class is unknown. A set of past observations (or a training set) of unknown classes is used to establish an empirical Bayes decision rule which performs like the Bayes rule and separates two classes with the probability of misclassification close to that of the Bayes rule. Monte Carlo simulation results are presented for several parametric distributions including normal and uniform distributions

Key words and phrases: classification, empirical Bayes, pattern recognittion.

---

author's address: Department of Mathematics, Rutgers University, Camden, NJ 08102. Tel:609-757-6439.

STATISTICAL MODELING OF A PRIORI INFORMATION FOR IMAGE PROCESSING PROBLEMS

Z. Liang

Dept. of Radiology, Duke University Medical Center, Durham, NC 27710

ABSTRACT

Statistical modeling of image processing problems of ill-posed in inverse process has been enhanced in recent years in terms of maximizing source entropy function (1-2) and in terms of maximizing data likelihood function (3-4). Although some effort has been made to consider both the source entropy and data likelihood information (5-6), statistical modeling of the image processing problems has not yet been extensively investigated. A formalism of Bayesian analysis incorporating the Poisson or Gaussian statistics of observed data accuratly is discussed in detail in this paper on different a priori source distribution probabilistic information. Most statistical methods can be derived from this formalism considering the different a priori source information. Systems of equations determining the Bayesian solutions were given for the different a priori source distribution information by maximizing the a posteriori probability given the observed data. Iterative Bayesian algorithms to carry out the calculation for the Bayesian solutions were derived using an expectation maximization technique (7). These algorithms were applied to computer simulated phantom imaging data. Improvement in image processing with these algorithms was demonstrated, compared to those algorithms of maximizing source entropy and data likelihood functions.

(1). R. Kikuchi and B. Soffer "Maximum Entropy Image Restoration I. The Entropy Expression" J. Opt. Soc. Am., vol.67, 1656-1665, 1977

(2). J. Skilling and S. Gull "The Entropy of an Image" SIAM Am. Math. Soc. Proc., vol.14, 169-189, 1984

(3). L. Shepp and Y. Vardi "Maximum Likelihood Reconstruction for Emission Tomography" IEEE Trans. Medical Imaging, vol.1, 113-122, 1982

(4). A. Rockmore and A. Macovski "A Maximum Likelihood Approach to Transmission Image Reconstruction from Projections" IEEE Trans. Nucl. Sci., vol. 24, 1929-1935, 1977

(5). B. Frieden "Restoring with Maximum Likelihood and Maximum Entropy" J. Opt. Soc. Am., vol.62, 511-518, 1972

(6). B. Frieden "Statistical Models for Image Restoration Problem" Computer Graph. Image Proc., vol.12, 40-58, 1980

(7). A. Demspter, N. Laird and D. Rubin "Maximum Likelihood from Incomplete Data via the EM algorithm" JRSS, vol.B, 1-38, 1977

A POOLED ERROR DENSITY ESTIMATE FOR THE BOOTSTRAP

Walter Liggett

National Bureau of Standards
Gaithersburg, MD 20899

Although a bootstrap based on resampling without replacement can be performed in the case of several small samples, a bootstrap based on a pooled density estimate is preferable if pooling is appropriate. In the case considered, the data consist of a few measurements on each of several dissimilar items, and the measurement errors are independent and identically distributed. The pooled error density estimate discussed is computed from first and second differences between measurements on the same item. Only first differences and therefore, only duplicate measurements, are needed if a symmetric error density is assumed. An error density that is possibly skewed requires triplicate measurements on some items. The error density estimate is based on the orthogonal expansion in Hermite functions and on the relation between the characteristic function of the error and the characteristic functions of the differences. A bootstrap based on this density estimate is applied in the case of items each measured three times. In this case, robust estimates of the item values can be computed. Several functions of the item values are potentially of interest. The range of the item values is considered. This is an interesting example because of the effect on this statistic of stretched-tailed error. Even with a robust estimator, the range of the item values is affected by stretched-tailed error because of the fact that robust estimators for samples of size three are not resistant to multiple contamination.

Computational aspect of harmonic signal detection

Keh-Shin Lii

Tai-Houn Tsou
Department of Statistics
University of California, Riverside

Detecting harmonic signal in a noisy enviroment is a classical problem and an important one. Typically, the noise process is assumed to be Gaussian. Therefore the analysis is mostly based upon second order theory such as covariance or periodogram. There are situations where the noise process is non-Gaussian then we can take advantage of the information contained in the higher order moments to possibly increase the efficiency of detecting the presence of harmonics.

This paper explores a method using both second order and higher order spectrum to ascertain the number of harmonics in the presence of non-white and non-Gaussian noise. Computational methods is discussed. Simulation examples are presented to indicate the effectiveness of the method in comparison with the classical second order methods.

IT'S TIME TO STOP

Hubert Lilliefors
George Washington University

This paper addresses the problem of determining the sample size to be used (when to stop sampling) when using a simulation to estimate the quantiles of the distribution of some statistic. Recently Dallal and Wilkinson (1986) used a procedure which started with a sample size of 50000 and computed a 95% confidence interval for the 99th quantile. If the width of the interval was less than some prescribed width (they used .001) they stopped. Otherwise they added another 50000 to the sample and tried again. This continued until either their condition was satisfied or they reached an upper limit on the sample size.

In this paper we present alternative procedures for determining when to stop the simulation which under certain circumstances may have some advantages over the Dallal and Wilkinson procedure. A simulation was used to compare the various procedures when estimating quantiles of several distributions.

For the alternative procedures, we make use of the well known asymptotic (normal) distribution of sample quantiles. Using this distribution it is straightforward to show that if we require a 95% probability that the sample quantile is within a distance B of the population quantile, then the sample size required is n=p(1-p)(1.96/B*f(x))**2, where x is the pth population quantile.We need an estimate for the density function evaluated at the population quantile.

Basically two estimators were used. These were the Siddiqui estimator (1960) and a new least squares estimator. We tried two basic procedures. 1.The first is a two stage procedure in which a preliminary sample was used to estimate the density function, which was used to calculate the required total sample size. From this the size of the additional sample size needed is determined. This second sample is drawn and the estimate for the quantile is determined using the two samples. 2. The second procedure is a three stage procedure in which, after the second sample is drawn,we again estimate the density function and if a larger sample is determined to be necessary we draw another sample.

The basic conclusion is that any of these procedures works reasonably well. Under certain circumstances our alternative procedures give improved results. In addition,they require stopping only once or twice to determine what additional sample size is needed. The Dallal and Wilkinson procedure will probably require many more such determinations.

References

Dallal, G.E. and Wilkinson, L. (1986), "An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality", *The American Statistician*, Vol 40, No.4, 294-296

Siddiqui, M. (1960), "Distribution of Quantiles in Samples from a Bivariate Population, *J. Res. Nat. Bur. of Standards*, B64, 145-150

A MODEL FOR INFORMATIVE CENSORING

William A. Link

Fish and Wildlife Service, Patuxent Wildlife Research Center, Laurel Md. 20708

Suppose that $T_1, T_2, \ldots, T_n$ is a random sample of "lifetimes" (non-negative continuous random variables) with common survival function $S(t) = P(T > t)$. We consider the problem of estimating $S(\cdot)$ when the T's are not directly observable; rather, one is able to observe $(X_1, \delta_1), (X_2, \delta_2), \ldots, (X_n, \delta_n)$ where $X_i \leq T_i$ and δ_i is a binary random variable equalling one if $X_i = T_i$ and zero otherwise.

The problem of estimating a survival function in the presence of random right censoring has been extensively studied. The majority of research has centered on the independent censoring model, in which $C_1, C_2, \ldots, C_n$ are "censoring times", independent of $T_1, T_2, \ldots T_n$, and $X_i = \min(T_i, C_i)$. Under this model, the Kaplan-Meier Estimator (KME) is the appropriate estimator of $S(\cdot)$.

It is not difficult to envision situations in which the assumption of independent censoring is inappropriate. However, if the only observations available are the pairs (X, δ), the independence assumption is completely untestable. It has been shown by Cox and Tsiatis that "there always exist independent censoring models consistent with any probability distribution for the observable pair (X, δ)" (Lagakos). The consequence of this is that if it is believed that the independence assumption is unwarranted, an equally untestable assumption about the joint distribution of (T, C) must be made.

If, however, covariates are observed in addition to (X, δ) the situation improves. We consider a model in which the population is divided into "high-risk" and "low-risk" subpopulations and in which censoring only occurs on lifetimes in the "high-risk" group. The "high-risk" subpopulation has hazard function $\lambda_H(\cdot) = m\lambda(\cdot)$, where $\lambda(\cdot)$ is the population hazard function and m is an unknown constant. Under this model, the KME yields substantial overestimates of S.

We consider an alternative estimation procedure in which the parameter m and the survival function are estimated by self-consistency algorithms.

Abstract

Brenda MacGibbon, Susan Groshen, Jean-Guy Levreault, Numerical Algorithms for Exact Calculations of Early Stopping Probabilities in One-Sample Clinical Trials with Censored Exponential Responses *

For some cancers, the existing treatment regimens produce long-term disease-free survival rates of 80% or better. In this situation a new protocol may aim to reduce the amount or duration of treatment, while maintaining the high disease-free survival rates. Although the primary goal is to evaluate the specific morbitity of such a new protocol, it is desirable to develop rules to stop the trial if many patients die or relapse early in the study and to study the statistical properties of these rules numerically. Since the failure (death or relapse) or success (survival) of the nth patient is not usually observed before the (n+1)st patient is entered onto the protocol, most developed sequential techniques do not apply to the problem. Most group sequential techniques involve large sample results, inappropriate for small studies. If the survival times of the patients follow an exponential distribution and the entry times into the trial are Poisson, and if these are independent, then a pure birth-and-death process with a well-defined transition matrix is an appropriate model. Analysis of the process enables the expression of error rates in terms of the transition probability matrix and renders these calculations computationally feasible. A conceptually simple design for monitoring a trial, in which a new treatment is evaluated after each observed failure, is presented and algorithms to calculate the error rates of interest are given. Algorithms for the calculation of the average sample number (ASN), the median and the quartiles of the sample size, as a function of the ratio of the entry rate to the failure rate, are constructed. Finally, the methods are illustrated on two examples involving the design of pilot studies.

---

* To be presented by Brenda MacGibbon, Department of Decision Sciences and Management Information Systems, Concordia University, Montreal, Canada.

A CLOSER LOOK AT SYMBOLIC COMPUTATION

William M. Makuch
John W. Wilkinson

Corporate Research and Development
General Electric Company
Schenectady, New York
and
Decisions Sciences & Engineering Systems Department
Rensselaer Polytechnic Institute
Troy, New York

ABSTRACT

For many, *symbolic computation,* is nothing more than a frustrating experience. The machine returns screen after screen of unmanageable expressions or fails on even the most simpliest of calculations. The typical novice user eventually questions the utility of a computer algebra approach. The problem here is generally not the capabilities of the symbolic system, nor is it the user's grandiose expectations. The problem is one of understanding the symbolic computation software and being able to successfully comunicate with it. This paper presents an initial exposure to some of the lesser known details which must be understood if the user intends on using symbolic systems beyond the elementary level.

An introductory level understanding of what a symbolic computation system can do is assumed. This paper then attempts to add a more complete understanding of symbolic representation, functional dependencies, evaluation, and simplification. The relevance of these topics to the computing statistician, as well as the strengths and limitations of computer algebra approaches, are also discussed. The MACSYMA system is used for illustrative purposes.

M

A NON-RANDOM WALK THROUGH FUTURES PRICES OF THE BRITISH POUND

William S. Mallios

California State University, Fresno

During 1984-86, foreign currencies reached record lows against the dollar, then recovered erratically. The period was characterized by high volatility and enormous losses. In such periods, currency modelling--for purposes of short term forecasting--would seem a natural recourse. However, results of such modelling appear infrequently in the literature. Possible reasons are (i) that random walk theory prevails (in reality or as a result of inadequate modelling) or (ii) that viable models are not publicized. Autoregressive-integrated- moving average (ARIMA) modelling, when applied to forecasting a particular currency without regard to relevant, contemporaneous variables, tends to support random walk theory. Such results are, however, misleading due to interrelations between leading currencies, precious metals, and their respective open interest.

To allow for such interrelations, a reduced system of equations is applied. Each dependent variable may be affected by its own lags and lagged shocks and/or those of other dependent variables, either in terms of first order or higher order modelling. Higher order terms include interactions between lagged variables. Analysis results for the British pound reject the random walk model and support the notion of second order modelling. Utilization of prior information in updating the model is presented in terms of empirical Bayes estimation.

RANDOM VARIABLES FOR SUPERCOMPUTERS

George Marsaglia

Department of Statistics
The Florida State University
Tallahassee, Florida 32303
(904-644-3218)
(marsagl@FSU)

A discussion of methods for generating random variables in supercomputers, particularly the 205 and ETA 10. Methods that exploit vector processing are well-suited for generating uniform random variables, both integer and real, and several of them are described. For non-uniform variates, however, methods that have proved best for conventional computers do not readily yield to vector methods. For example, the best methods for normal or exponential variates in conventional computers take less than 1.2 T, where T is the time for a uniform variate, yet in supercomputers those methods take relatively much longer. Different approaches to reducing these times will be discussed.

M

Maximum Queue Size and Hashing with Lazy Deletion

Claire M. Mathieu[1] and *Jeffrey Scott Vitter*[2]

Abstract. We answer questions about the distribution of the maximum size of queues and sweepline processes. Queuing phenomena are widespread in the fields of operating systems, distributed systems, and performance evaluation. Queues also arise directly as constructs in computer programs, for example, in the form of sweepline data structures for geometric applications, buffers, dictionaries, sets, stacks, queues, and priority queues. The concept of "maximum" occurs in many issues of resource allocation. If the size of the queue represents the amount of resource used by a computer program or a systems component, then such information is important making intelligent decisions about preallocating resources.

In this paper we study general birth-and-death processes, the $M/G/\infty$ model, and a non-Markovian process (algorithm) for processing plane sweepline information, called ***hashing with lazy deletion*** (HwLD), introduced recently by Vitter and Van Wyk in ***Algorithmica***. It has been shown that HwLD is optimal in terms of expected time and dynamic space, up to a constant factor; our results show that it is also optimal in terms of expected ***preallocated*** space. Our results also show strong links between the maximum sizes of continuous phenomena and of their discrete counterparts.

We obtain an array of results about the maximum queue size using two independent approaches. In our first approach, we develop several formulas for the distribution of the maximum queue size for general birth-and-death processes (which includes the $M/M/\infty$ process) and HwLD. The formulas provide exact numerical data on the distributions, and in some cases lead to asymptotics as the time interval grows. There is a common underlying structure in the formulas for the different models: the transform of interest in each case is the ratio of consecutive classical orthogonal polynomials. And the particular polynomials involved give a strong link to the maximum size of file histories, as studied combinatorically by Flajolet, Françon, and Vuillemin.

In our second approach, we get optimal big-oh bounds on the expected maximum queue size in the general $M/G/\infty$ model (which includes $M/M/\infty$ as a special case) by using non-queueing theory techniques from the analysis of algorithms. We approximate the maximum queue size (and, in the case of HwLD, also the maximum data structure size) in a novel way by sums of discrete quantities related to hashing, specifically, maximum slot occupancies. (The hashing in our approximation scheme has nothing to do with the hashing inherent in HwLD.) Our techniques also seem applicable to other queueing models, such as $M/M/1$.

---

[1] Current address: Laboratoire d'Informatique de l'Ecole Normale Supérieure, 45, rue d'Ulm, 75230 Paris Cedex 05, France. Research was also done while the author was at Princeton University and funded by a Proctor Fellowship.

[2] Current address: Department of Computer Science, Brown University, Box 1910, Providence, R. I. 02912. Research was also done while the author was on sabbatical at Ecole Normale Supérieure and INRIA. Support was provided in part by NSF research grant DCR-84-03613, by an NSF Presidential Young Investigator Award with matching funds from an IBM Faculty Development Award and an AT&T research grant, and by a Guggenheim Fellowship.

INTRODUCTION TO PACKET-SWITCHING NETWORKS

Jeffrey Mayersohn
BBN Communications Corporation
50 Moulton Street
Cambridge, Massachusetts 02238
(Electronic Mail: mayersoh@ALEXANDER.BBN.COM)

ABSTRACT

Communications networks allow transmission resources to be shared by a large population of users. Packet switching is a particular type of network technology in which the data to be transmitted are divided into discrete units, called packets. These packets independently travel from the source to the destination, where they are reassembled into their original form. Among the mathematical problems associated with packet-switching networks are the design of optimal network configurations and the development of network control algorithms. An example of the latter type of algorithm is routing, which determines the path that will be taken by each packet through the network. Another class of problems concerns the analysis of network performance. Packet switching will be discussed and examples of solutions to the above problems will be discussed within the context of the ARPANET, which was one of the first packet-switching networks.

Application of Posterior Approximation Techniques to the Ordered Dirichlet Distribution

Thomas A. Mazzuchi
Refik Soyer

George Washington University

The ordered Dirichlet distribution has been shown to be a meaningful prior distribution for the analysis of several important problems in reliability and biometry. Unfortunately, the relevant posterior quantities can rarely be obtained in simple closed form. Closed form results that are obtained are often complex and subject to numerical error due to their dependence on the extreme range of the gamma function. Often numerical error and computation time increase with the sample size. In this paper we explore the use of a posterior approximation technique recently suggested by Tierney and Kadane (1986) in these cases. We thus illustrate a multivariate application of these techniques as well as a comparison of the accuracy of these approximation techniques with the closed form solution.

M

COMBINING KNOWLEDGE ACQUISITION AND CLASSICAL STATISTICAL TECHNIQUES IN THE DEVELOPMENT OF A VETERINARY MEDICAL EXPERT SYSTEM

Dr. Mary McLeish
Departments of Computing and Information Science/Statistics
University of Guelph
Guelph, Ontario
N1G 2W1

A project was recently begun as the University of Guelph between the departments of Computing Science, Statistics and the Ontario College of Veterinary Medicine. Equine colic was chosen for a prototype domain, due to the diagnostic difficulty of predicting true surgical cases. Unnecessary surgeries are costly and can have long term debilitating effects on a productive animal. A sophisticated medical information system at OVC has been in operation for 10 years and has collected a vast amount of on-line medical data. Many test results are fed automatically into the database. It was our intent to design a system which was largely driven by rules and information extracted from this enormous statistical source.

The role of probability and statistics in the development of expert systems is discussed in books such as "Artificial Intelligence and Statistics", by W.A. Gale. The methodologies employed by the early large medical projects, like MYCIN (Stanford), often used ad-hoc factors to combine uncertain information and were concerned primarily with imitating the mental reasoning processes of doctors. In a recent paper by Drs. Patil, Schwartz and Szolovits in the New England Journal of Medicine (Vol. 16, 1987) it is suggested that it is time to link the old with the new -the old being classical statistical routines, such as discriminant analysis. To quote, "now that much of the A.I. community has turned to causal, pathophysiologic reasoning, it has become apparent that some of the earlier, discarded diagnostic strategies may have important value in enhancing the performance of new programs ..." To successfully merge the different available approaches is a difficult one, which these authors recognize when they state that "an extensive research effort is required before all these techniques can be incorporated into a single program".

The project at hand is using a variety of data analysis techniques, uncertainty management tools and human expertise to build the type of system just suggested. Discriminant analysis techniques were tried on data sets involving 45 input parameters in two groups: clinical data, such as pain, temperature, pulse, results of rectal examinations, and pathology data: total cell counts, protein levels, etc. The most significant variables were two very subjective measures: pain and abdominal distension. The pathology data did not seem to influence the decision process. The decision tree obtained produced a tendency to over-operate.

In an attempt to discover other relevant parameters and not discount the pathology data a number of other knowledge acquisition techniques not assuming linearity or normality of variables were tried on the same data. These included an event-covering method (Dr.Chiu and A. Wong, Pattern Analysis group, U. of Waterloo), an inductive learning technique (Dr. L. Rendell, University of Illinois, Urbana Champaign) and the learning (max entropy) approach of R. Quinlin (University of Sydney). These routines did discover other significant factors in the clinical data and interesting relationships between variables (clusters). They also discovered significant factors in the pathology data. Some of these methodologies were less sensitive to missing data than statistical routines, like discriminant analysis. With some methods, missing data was a very serious problem. As we were not doing analysis to strictly publish the statistical results, but to aid us with over-all diagnostic strategy, we constructed new data sets with estimated missing values. Logistic regression was run on the new data sets to compare results with the earlier discriminant analysis and this generally gave more informative results.

Other techniques being tried include a Bayesian inductive technique due to Peter Cheeseman. This provides interesting data classifications not dependent on any form of similarity measure (distance etc.). These results may be used in a predictive manner e.g. by noting the occurrence of surgeries in a class and using this as an indicator for an incoming case found to belong to that class.

The above mentioned methods usually discard variables of low predictive power. The uncertainty management techniques, often used in expert systems, include all symptoms and provide mechanisms for combination of evidence. Bayesian approach, Dempster-Shafer theory, etc). We are now implementing a 'fuzzy approach (using fuzzy relations) somewhat like that used in the CARDIAG system in Austria. This is partly to test whether methods working with very few variables are as useful for diagnostic purposes as methods including all possible symptoms.

We are now undertaking the difficult task of integrating results from these various methods with medical expertise to build an on-line system and test it on incoming cases. The full paper will describe the methodologies and results in more detail along with the design of the expert system.

SMOOTHING IRREGULAR TIME SERIES

by

A. Ian McLeod
Dept. of Statistical and Actuarial Sciences
The University of Western Ontario
London, Ontario, Canada, N6A 5B9

Keith W. Hipel
Departments of Systems Design Engineering
and Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1

and

Byron Bodo
Water Resources Branch
Ministry of the Environment
135 St. Clair Avenue West
Toronto, Ontario, Canada, M4V 1P5

ABSTRACT

In 1979, Cleveland introduced the method of robust locally weighted regression for smoothing data (x_t, y_t), $t=1,\ldots,n$. This method is extended to handle irregularly spaced-seasonal time series. The smoothed value for the rth year and mth month is represented as

$$\hat{z}_{r,m} = \hat{\mu}_t - \hat{\alpha}_r r - \hat{\beta}_m m - \hat{\gamma}_{r,m} t$$

where $\hat{\mu}_t$, $\hat{\alpha}_r$, $\hat{\beta}_m$ and $\hat{\gamma}_{r,m}$ are determined by robust locally weighted least squares. Efficient APL programs for implementing this procedure are developed. Tests for the absence of moving seasonality $(H_0: \gamma_{r,m}=0)$ and for the absence of trends $(H_0: \alpha_r = \gamma_{r,m} = 0)$ are developed by bootstrapping the regression. The usefulness of the new methodology for interpreting environmental water quality parameters is discussed.

SIMULATED ANNEALING IN THE CONSTRUCTION OF OPTIMAL DESIGN

Ruth K. Meyer
Business Computer Information Systems
St. Cloud State University
St. Cloud, Minnesota

Christopher J. Nachtsheim
Department of Management Sciences
University of Minnesota
Minneapolis, Minnesota

ABSTRACT

Exact optimal designs have generally been constructed using a finite design space and various exchange algorithms, which oftentimes converge at a local optimum. Branch-and-bound methods guarantee optimal designs, but are computationally infeasible for large problems. We apply the generalized simulated annealing algorithm to the construction of exact optimal designs on both finite and continuous design spaces, and evaluate its effectiveness. We present optimal designs for large dimensional problems.

Minimum Cost Path Planning in the Random Traversability Space

P. Graglia, A. Meystel
Drexel University
Philadelphia, PA 19104

Abstract

Random traversability space (RT-space) is introduced and developed as the most general spatial representation for the path planning system of autonomous robots. It is demonstrated that any physical spatial situation can be mapped into RT-space, and the quantitative model can be built using the statistical characteristics of the physical spatial situation. A mathematical abstract model of autonomous robot is explored which is understood as a dimensionless stochastic automaton pursuing a goal while modifying its behavior as new information is acquired about its random spatial environment. A formalism for the automaton is proposed linking the stochastic input with the description of the automaton vicinity, and the deterministic output with the motion of the robot-automaton. The flow of information through the system should provide for minimum cost motion of the robot-automaton toward the goal.

The computation model of the robot-automaton is of interest. A second (generalized) level of traversability space is introduced to reduce computational complexity and make tractable the problem of stochastic minimum cost control. The generalized level of representation is used to guide search in the original RT-space. A theorem is proven concerned with the assignment of the minimal bounds of the search envelope. It is shown that the process of generalization affects the statistical characteristics of the search space. Comparisons are made between the results of the robot-automaton operation with different envelopes of search and under different heuristics of search.

A process of recursive generalization is considered in the RT-space which leads to the hierarchical RT-representation, and to the subsequent recursive hierarchical algorithm of computation. This is done with successively smaller envelopes of search and the results are analyzed with respect to relative error from the optimal path. The system is intended to develop joint hierarchical planning/control sequences based both upon the knowledge stored in the memory and/or acquired during the robot-automaton operation. The path planning system combines the spatial map of the vicinity and spatial knowledge about the larger subset of the environment including the final goal, to form a complete state description of the system. A goal-oriented procedure of path planning is then applied which generates a sequence of states which best satisfies the condition of minimum cost goal goal achievement and is considered the path. A variety of simulation experiments is considered for different traversability spaces. The results of comparison are given with the conventional algorithms of dealing with the problem.

UNBIASED ESTIMATES OF MULTIVARIATE GENERAL MOMENT FUNCTIONS OF THE POPULATION AND APPLICATION TO SAMPLING WITHOUT REPLACEMENT FROM A FINITE POPULATION

by

N.N. Mikhail
Liberty University

Abstract

Unbiased estimates of the multivariate general moment functions of the population are obtained when sampling from finite populations. Partitions and power sums are featured. Unbiased estimates of multivariate cumulants and moment functions are obtained as examples of application.

Symposium on the interface: Computer Science and Statistics

ABSTRACT

P. Warwick Millar
Univ. of Calif.
Berkeley, CA 94720

STOCHASTIC TEST STATISTICS

Stochastic procedures are tests, estimates, or confidence sets which have two properties: (a) they are functions of the data sample plus an auxiliary random sample (b) they become nearly non-randomized as the sample sizes increase. Such procedures arise as numerically feasible, computationally intensive, approximations to numerically intractable procedures. They often involve iterated bootstrap techniques together with random searches over abstract populations.

Let $\{P_\theta, \theta \in \Theta\}$ be a family of probabilities on R^d. A plausible test statistic for the null hypothesis that the correct model is $\{P_\theta, \theta \in \Theta\}$ might be: $G_n \equiv \inf_\theta \sup_A n^{1/2}|P_n(A) - P_\theta(A)|$ where the sup is over all half-spaces in R^d and P_n is the empirical measure. Of course the null hypothesis would be rejected for large values of G_n. In most cases of interest, where $d \geq 2$ and Θ is "infinite dimensional", (i.e., nonparametric) the statistic G_n is virtually uncomputable. A related <u>stochastic</u> goodness of fit statistic with attractive asymptotic properties consists in (a) replacing the inf in the definition of G_n by a minimum over a random collection of θ's, consisting of j_n bootstrap replicas of a preliminary $n^{1/2}$ consistent estimator of θ, and (b) replacing the sup by a maximum over k_n sets chosen at random. Val critical values can then be obtained by bootstrap applied to this (computationally feasible) stochastic GOF statistic. These stochastic GOF statistics have been analysed in detail for two particular nonparametric models P_θ : location models on R^d, $d \geq 2$, and the logistic model.

This talk surveys some of these recent results.

BOOTSTRAP PROCEDURES IN RANDOM EFFECT MODELS F PARING RESPONSE RATES IN MULTI-CENTER CLINICAL

Michael F. Miller, Ph.D.
Hoechst-Roussel Pharmaceuticals Inc
Somerville, N.J. 08876

Let <P(j), Q(j)>, j=1,2,--k be population plac treatment response rates (probabilities) at each of k centers lti-center clinical trial. Let L(j) = <LP(j), LQ(j)> be the c ding logits (ln(P/(1-P)) of P(j), Q(j) respectively. In this st L(j)'s are assumed to be random vectors, i.i.d., having common p.d.f. g. Letting gp, gq denote the marginal p.d.f.'s of LP, L en the no treatment effect null hypothesis proposed here is g, The estimated logits from placebo and treatment patients at each ce are given by LH(j) = <LHP(j), LHQ(j)>, j=1,2,--k. Conditioned on , the distribution of LH(j) is approximately bivariate normal with mean L(j) and diagonal covariance matrix D_j containing the estimated variances of the estimated logits. Based on the observed LH(j)'s, estimates of the joint p.d.f. g, and hence gp, gq, will be investigated. Appropriate functionals of these estimates will be used to compare gp and gq. The sampling distributions of these functionals (means, weighted percentiles) will be studied using a two stage bootstrap simulation: generate population logits from the estimate of g, then generate success/failure data for each center conditioned on these population logits. A discussion of the computer implementation of this methodology will be presented along with an analysis of real clinical trial data.

Title: Computation of the Theoretical Autocovariance Function of Multivariate ARMA Processes

Author: Stefan Mittnik

Address: Department of Economics
SUNY at Stony Brook
Stony Brook, NY 11974-4384

Phone: 516-632-7532

Abstract

The theoretical autocovariance function is an important instrument in time series modelling. The derivation of the exact likelihood function of ARMA models requires the specification of the theoretical autocovariances in terms of the model parameters. The autocovariance function plays also a crucial role in model identification procedures. Nicholls & Hall (1979) provide a closed form expression for the theoretical autocovariances of multivariate ARMA models. Ansley (1980) and Kohn & Ansley (1982) present rather complex algorithms which are computationally more efficient than the one in Nicholls & Hall (1979).

Here we suggest simpler closed form expressions that provide more insight into the relationship of autocovariances and ARMA parameters. They are particularly useful when estimating moving average parameters via factorization methods and in evaluating the exact maximum likelihood function of ARMA models. The results enable us to compare the algorithms of Nicholls & Hall (1979), Ansley (1980) and Kohn & Ansley (1982) by fitting them into a general framework.

References

ANSLEY, C. F. (1980). Computation of the theoretical autocovariance function for a vector ARMA process. *J. Statist. Comput. & Simul.* 12, 15-24.

KOHN, R. & ANSLEY C. F. (1982). A note on obtaining the theoretical autocovariances of an ARMA process. *J. Statist. Comput. & Simul.* 15, 273-83.

NICHOLLS, D. F. & A. D. HALL (1979). The exact likelihood function of multivariate autoregressive-moving average models. *Biometrika* 66, 259-64.

M

Abstract On Increasing Reliability of Multiversion Fault-Tolerant Software Design by Modularization

Junryo Miyashita
Department of Computer Science
California State University at San Bernardino

Fault-tolerant software achieves its fault-tolerance by introducing redundancies in software. Well known Fault-Tolerant software designs are: 1) N-version programmings, 2) Recovery Block, and 3) Consensus Recovery Block. These designs all use several versions of a program to achieve their reliabilities. They shall be refered to as "multiversion fault-tolerant software design". One problem of developing multi-versions of a program is the high cost of development. This paper addresses that problem. Rather than working on the common requirement specification for a whole program, teams of programmers will work on the common specifications for each module in a program. A program consists of a set of modules. This will enable the modules in each version to be interchangeable.

Theoretical reliabilities of modularized multi-version fault-tolerant software are derived in closed forms. The numerical results of the modularization effects on the reliabilities on the three well known multi-version fault-tolerant software are calculated and the complete results are given in table forms. The numerical results show the dramatic increase in reliabilities in each multiversion softwares. For example:

In N-version programming, Assume R(i,j) = R *for* all *i and j*. That is, the reliabilities of each module of each version is constantly R then for example when R = .90 , 3 (i.e.n=3)original versions and each version has 2 parts(modules:i.e.m=2) then the modularization will increase the reliability of the software by 1.7 times compared to N-versions without modularization. When n = 4 and m = 3 then the increase in the reliability is 5.7. If n = 5 and m = 8 then the increase is about 77 times. If R = 9.8 and n = 5 and m =8 then the increase in the reliablity is about 327 times. So the numerical results indicate that by modularization any increase in number of original versions or increase in the number of modules will increase the reliability of the software in significant amounts.

In Recovery Block, the reliability of the software depends on the reliability of versions as well as the reliability of the acceptance test. *If* the reliability *of* the acceptance test is low, then no increase in the reliability *of the* versions can increase the reliability of the software much. Assumming that the acceptance test reliability is very high or perfect, then the modularization will increase the reliability of the software more significantly than that of N-version programming. Results to this effect will be given in the tables.

Consensus Recovery Block overcomes the weakness of the Recovery Block by eliminating the heavy dependencies on the acceptance test by first doing N-version programming.It also eliminates the weakness of N-version programming on non-agreements by incorporating the acceptance test in case of non-agreements. The increase in the reliability is more significant than either N-version programming or Recovery Block schemes if the acceptance test is near perfect. Even if the acceptance test reliability is rather low, it still does significantly better than Consensus Recovery Block without modularizations.

ALGORITHMS TO RECONSTRUCT A CONVEX SET FROM SAMPLE POINTS

M. Moore, École Polytechnique and McGill University
Y. Lemay, Bell Canada
S. Archambault, École Polytechnique

Let C be an unknown compact convex set in the plane and suppose the sample points, $X_1, \dots X_n$ are selected independently according to a distribution function F on R^2 whose support includes C. For each sample point, in addition to its coordinates it is known if it is interior or exterior to C. Based on this information it is desired to reconstruct (estimate) C. A similar problem, where only uniform sample points on C are observed, has been considered by Ripley and Rasson (J. App. Prob., 14, 483-491) and Moore (Ann. Statist., 12, 1090- 1100).

The sample space is made of the vectors $(x_1, i_1, \dots, x_n, i_n)$ where x_j represents the coordinates of the jth sample point, $i_j = 1$ if this sample point is in C and $i_j = 0$ otherwise, $j = 1, \dots, n$. Let H denotes the convex hull of the sample points x_j for which $i_j = 1$ (interior points) and let

$$K = \bigcup_{j \in E} \{x: x = x_j + \lambda(x_j - y), \ y \in H, \ \lambda \geq 0\}$$

where $E = \{j: i_j = 0\}$. The unknown convex set C includes H and is included in the complement of K. Let V be the set of vertices of H and T be the set of peaks of K (a peak is a sample point outside C whose removal would change K). It can be shown that the pair (V,T) is a minimal sufficient statistic for the family $\{P_C; C \in \mathcal{C}\}$, $\mathcal{C}$ being the class of compact convex sets in the plane and P_C is the probability measure on the sample space given C and the distribution F. A natural criteri⌄ to evaluate a reconstruction rule δ is

$$(1) \qquad R[C,\delta] = E[m(C \,\Delta\, \delta(x_1, i_1 \dots, x_n, i_n))],$$

m denoting the Lebesge measure and the expectation being with respect to P_C. It seems difficult to obtain a procedure δ^* based on (V,T) which is in some sense optimal with respect to (1) (e.g. mimimax).

In this paper we propose three algorithms to reconstruct C. In increasing complexity order these reconstructions are:

a) a dilation of H by a unique factor determined by V,

b) a deformation of H obtained by applying a particular dilation factor to each side of H; these dilation factors being determined by the appropriate elements of V.

c) an average (Minkowski addition) of two reconstructions, the first being simply H and the second being obtained mainly from V.

By a simulation experiment these algorithms are compared using a criteria related to (1). The algorithm c) is quite complex and requires much geometrical computations, but presents definite advantages in regard to precision and stability.

Block Truncated-Newton Methods for Parallel Optimization

Stephen G. Nash

Ariela Sofer

Operations Research and Applied Statistics Department
George Mason University
Fairfax, VA 22030

ABSTRACT

Truncated-Newton methods are a class of optimization methods suitable for large-scale problems. At each iteration, a search direction is obtained by approximately solving the Newton equations using an iterative method. In this way, matrix costs and second-derivative calculations are avoided, hence removing the major drawbacks of Newton's method. In this form, the algorithms are well-suited for vectorization. Further improvements in performance are sought by using block iterative methods for computing the search direction. In particular, conjugate-gradient-type methods are considered. Computational experience on a hypercube computer will be reported.

An Example of the Use of a Bayesian Interpretation of Multiple Discriminant Analysis Results

James R. Nolan
Siena College

The use of Bayesian statistics to add additional information about the results of a binary dependent variable multiple discriminant analysis will be detailed using a recently completed study.

Several methods are examined for determining the best discriminant function, e.g. Wilks' Lambda, eigenvalues, canonical correlation. The usual procedure is to then examine the "confusion matrix" and draw conclusions about the predictive power of the discriminant function. Far more information can be obtained by employing Bayesian statistics to examine, for any actual or hypothetical case, the probability of obtaining a particular value of the binary dependent variable. Thus one can obtain, on a case by case basis, a measure of the "strength" of the discriminant function predicted value of the discrete dependent variable.

Details about the computer statistical software package utilized for this analysis will be given and several pages of output will be available in the form of handouts.

N

Comparison of "Local Model" Classification Methods

Daniel Normolle

Department of Biostatistics

University of Michigan

A large Monte Carlo study is reported that compares three "local" methods (Classification by Kernel PDF Estimation, Cross-Validated Nearest-Neighbor Classification, and Tree Classification with Pruning for Optimality), a benchmark method (Bayes' Classification Rule), and three "global" methods (Linear Discriminant Analysis, Logistic Regression, and Quadratic Discriminant Function on Normal Scores) with respect to their ability to correctly classify test samples.

The data are drawn from a 5x2x2x2x3 completely crossed design, where the levels of analysis are Distribution Type (Gaussian, Cauchy, Lognormal, Bimodal, Uniform), Dimension (2, 6), Class-Conditional Dispersion (Equal, Unequal), Separation of Classes (Low, High), and Training Sample Size (40, 80, 160). Each design cell is replicated 100 times, yielding a total of 84,000 classification runs. Thus, the experiment compares three local methods, each with an associated optimizing procedure, on level ground over a wide variety of data situations. The results of the experiment are described using statistical techniques (e.g., MANOVA) and graphical techniques, such as Andrew's curves.

The nearest-neighbor and classification tree methods are found to be roughly equivalent, with the nearest-neighbor preferable on well-separated data, and the classification tree better with larger sample sizes. PDF Estimation is superior to the other two local model methods on the two-dimensional data, but weakens considerably on the six-dimensional data. The three local model methods are superior to the ordinary Linear Discriminant Function on non-Gaussian data, but are bested by the use of the Quadratic Discriminant Function on the Normal Scores almost uniformly.

Mice, rain forests and finches: experiences collaborating with biologists

Doug Nychka
North Carolina State University
Department of Statistics

In the first part of this talk I would like to discuss some of my experiences working with biologists in cancer research, tropical ecology and population genetics. Besides describing some of the new statistics that have been developed, the role of computing in these projects will also be stressed. With the proliferation of microcomputers, researchers are often able to collect novel experimental data. It is a challenge to statisticians to develop the tools to alalyze these more complex experimental results. The second part of this talk will give some details about using projection pursuit techniques for estimating fitness surfaces in population genetics. When the smoothness of the ridge functions is chosen adaptively by cross validation, projection pursuit becomes a computationally intensive technique. As an example, the overwinter survival of song sparrows is related to various morphological measurements. This relationship is important because it may suggest what characteristics are being favored through natural selection.

Image Analysis of the Microvascular System in the Rat Cremaster Muscle

by

C. O'Connor, P. D. Harris, A. Desoky, and G. Ighodaro

A VAX-based image processing system has been developed for the digitization and analysis of the microvascular system in the rat cremaster muscle. These are images of blood vessels which are less than one millimeter in diameter. The purpose of this system is to obtain quantitative morphometric data on the microvascular system which cannot be easily obtained by manual methods. Animal studies have shown that microcirculation can be used in the detection of certain systemic vascular disceases such as diabetes mellitus and hypertension. These diseases involve major disturbances in the dimensions and the distributions of microvessels. The developed techniques are being used to determine the blood vessel distributions for a number of samples. Statistical testing will be made on samples of images comprising diseased and nondiseased animals, to determine which image component parameters best discriminate diseased and nondiseased samples.

Statistical Computing on a Hypercube

George Ostrouchov, Oak Ridge National Laboratory

A hypercube parallel computer is a network of 2^n processors, each with only local memory, whose activities are coordinated by messages the processors send between themselves. The interconnection network corresponds to the edges of an n-dimensional cube with a processor at each vertex. Some recent experiences and results in developing a hypercube algorithm for iterative proportional fitting of large Poisson regression problems will be discussed. The algorithm is implemented on a 64-processor INTEL iPSC hypercube.

Empirical Likelihood Confidence Regions

Art Owen
Dpartment of Statistics
Stanford University

An empirical likelihood ratio function is identified and used to obtain confidence regions for vector valued statistical functionals. The result is a nonparametric version of Wilks' (1938) theorem and a multivariate generalization of Owen (1987). Cornish-Fisher expansions show that the empirical likelihood intervals for a one dimensional mean are less adversely affected by skewness that are those based on student's t statistic. An effective computational strategy is presented for maximising the empirical likelihood ratio function. The main tool is a dual problem of smaller dimension for which there are algorithms that converge to the unique global solution from any starting point. The technique is used to justify nonparametric intervals for variances, correlations and regression parameters.

Newton Methods for B-Differentiable Functions: Theory and Applications

Jong-Shi Pang

Mathematical Sciences Department
The Johns Hopkins University
Baltimore, MD 21218

ABSTRACT

In this paper, we extend the classical Newton method for solving systems of nonlinear equations to the class of problems with B-differentiable functions. Such functions were defined by S.M. Robinson and possess differentiability properties weaker than Frechet-differentiability. We demonstrate that all the basic convergence properties of the classical Newton method and its many modifications are preserved in the extension. We discuss applications of the results to many problems in mathematical programming. These applications lead to interesting second-order active-set-Newton-combined methods for solving the problems discussed.

Abstract for:

An Approximate Confidence Interval for the Optimal Number of Mammography X-ray Units in the Dallas-Forth Worth Metropolitan Area

Roger Peck, University of Rhode Island

The American Cancer Society was interested in geographically locating mammography x-ray units in a five county area of the Dallas-Fort Worth metropolitan area based on 1980 census tract data consisting of the x,y co-ordinant location (adjusted to reflect real distance) of 376,256 women aged 35 to 65. We decided to determine an approximate confidence interval for the number of units that would be needed to insure proper coverage of the area and yet be cost effective.

This is a clustering problem in which the optimal number of clusters (the number of units that the area can support) needs to be determined along with their respective cluster centers (the locations of the units). The quality of any clustering is measured by a loss function which takes into account both the cost of operating the units and the cost associated with the likelihood of a woman not using one of the units. Peck, Van Ness, and Fisher (1988) have shown that a "best" clustering can be obtained by minimizing this loss function. They have also developed a bootstrap-based procedure for obtaining approximate confidence bounds on the number of clusters in the "best" clustering.

In this problem, the two cost functions can easily be determined. The first cost function can be determined from the fact that the units cost approximately $300,000 for startup and $100,000 per year for personnel and maintenance. It can be argued, that the other cost function is a function of the distance a woman lives from a unit, that is, women living near to a unit are more likely to use it than women living further away. Given the cost functions and the census tract data the approximate confidence interval for the optimal number of units can be determined along with their corresponding cluster centers.

Key Words: Cluster Analysis; K-means Clustering; Bootstrap; Confidence Interval; Simulation Study.

P

Statistical Methods for Document Retrieval and Browsing

Jan Pedersen, Xerox PARC, J.W. Tukey and P.K. Halvorse

I will discuss the interaction between statistics and the vision of document retrieval and browsing currently being developed at Xerox PARC as part of a research initiative examining the implications of the "paperless office". Given that filing of extremely large volumes of textual and graphical information will soon be feasible, if it is not already so, the problem of "unfiling" will assume greater importance.

The PARC vision of retrieval favors high band-width interaction with the user rather than the traditional emphasis on query languages. It is thought that the combination of certain aspects of computational linguistics to extract a meaningful summary of the content of a document and interactive subset selection will out perform traditional keyword based queries. I will discuss one such retrieval and browsing technique based on content word triples.

Estimation of the variance matrix for maximum likelihood parameters by quasi-Newton methods

Linda Williams Pickle
National Cancer Institute

Garth P. McCormick
George Washington University

Much work has been done to develop methods for solving unconstrained optimization problems that do not require specification of second derivatives of the objective function, which can be extremely complex. While the rate of convergence of these quasi-Newton methods to the correct solution vector has been shown to be superlinear, little research has been done on the behavior of the convergence of the inverse Hessian approximation to its true value. These optimization methods are now being used in new microcomputer statistical packages to calculate maximum likelihood parameter estimates, and the resulting inverse Hessian matrix is being used as an asymptotic variance estimator for the parameters. We have examined the behavior of this matrix approximation for several representative problems. Comparison of known analytic results to results from the BFGS quasi-Newton method using an optimal step size suggests that after the first n iterations (n = number of parameters to be estimated) the matrix approximation then converges at about the same rate as the parameter vector. We examine several functions useful as candidates for additional convergence criteria to ensure accuracy of the variance matrix approximation in practice or to identify situations where the approximation might be poor.

ABSTRACT

Exact Power Calculation for the Chi-Square Test of Two Proportions

Carl E. Pierchala
Food and Drug Administration

In calculating the power of the Pearson Chi-Square test of two independent proportions, it is usual to use an approximation. This can speed up the computations and simplify programming. At times, however, it is useful to directly compute the exact power. For example, one may wish to assess an approximation's adequacy in a specific situation. Thus, an APL program was developed to do exact power calculations on an IBM PC/XT. It gives accurate and reasonably fast computations. The exact power values for certain circumstances are compared to the corresponding values obtained using an approximation based on the arc sine transformation. It is shown that this approximation is quite inaccurate in some situations. Also, the program is used to demonstrate that the exact size of the test can differ dramatically from the nominal size.

Bootstrapping the Mixed Regression Model with Reference to the Capital and Energy Complementarity Debate*

Baldev Raj
Wilfrid Laurier University

ABSTRACT

This study empirically investigates the usefulness of bootstrapping the standard error of estimates of the Hicks-Allen elasticity of substitution (AES) as obtained from the Mixed Regression model, with specific reference to the capital-energy complementary debate. This is accomplished by obtaining the bootstrap standard error of estimate of the AES for capital and energy in the cost-share equations when homogeneity and symmetry constraints are imposed stochastically over 500 simulation runs as opposed to deterministically, which earlier studies assumed. Our results show that the bootstrap provides an accurate method of obtaining the standard error of estimate (SEOE) of the AES while the asymptotic formula can overestimate the small sample SEOE by over 70 percent. Based on interval estimates of the AES for capital and energy the bootstrap SEOE cannot reject the substitutability hypothesis even though the point estimate does support the complementarity hypothesis. The data generating processes used in the simulations are based on previous studies by Berndt and Wood (1975, 1979), among others.

ABSTRACT

Classifying linear mixtures with an application to high resolution gas chromatography

William S. Rayens
University of Kentucky

This paper proposes an elegant, yet straightforward model for classifying linear mixtures. A linear mixture is defined as a random vector y in which the variables are a (nonnegative) weighted average of corresponding variables, assumed to characterize g component groups. These weights are referred to as "mixing proportions". The model seeks to identify the mixture constituents and estimate the mixing proportions. It is demonstrated within the context of high resolution gas chromatography and the problem of identifying the constituents in polychlorinated biphenyl mixtures.

Structure and Finiteness Conditions on Graphs

Neil Robertson
Department of Mathematics
Ohio State University

Graphs are finite objects consisting of two sets, a vertex-set and an edge-set; where each edge is associated with two (not necessarily distinct) vertices. Such objects are ubiquitous in the real world and lend themselves readily to algorithmic questions concerning certain structural properties they may or may not possess. Through joint work with Paul Seymour of Bell Communications Research over the past six years a very extensive theory has been developed of certain types of graph structures studied in combinatorial optimization. Three closely related kinds of theorems have resulted; (1) structure theorems for which, if a graph does not have a certain type of internal structure then it possesses an external structure of a certain type, (2) finiteness theorems which say that for a given external structure there is a finite number of minimal graphs not possessing that structure (obstacles), and (3) algorithms, running in polynomial time, which given any finite graph and any fixed structure type either exhibit the structure on the graph or an obstacle to the structure within the graph. These algorithms are developments of results dating back up to sixty years and answer several longstanding open questions. They also have some unusual features of interest to the general theory of algorithms which has been developed so extensively in recent years.

On The Probability Integrals Of The Multivariate Normal; The 2^n-Tree and The Monte-Carlo Techniques.

Dror Rom and Sanat Sarkar

Department Of Statistics, Temple University, Philadelphia, Pennsylvania.

Abstract

Two techniques are proposed for computing probability integrals of the multivariate normal distribution. The first technique is based on the 2^n-tree scheme and is shown to perform well even for the near singular distribution. The technique employs a tree structure to represent the multivariate density. This representation gives a fast and efficient partition of the n-space and in general requires substantially less computations than other available techniques.

The second technique is essentially a variance reduction improvement of the Monte-Carlo integration method. As a technique based on simulation the Monte-Carlo method suffers from random variability, however it is still a usefull approach when the dimensionality is high. The proposed technique is shown to reduce the variance of the Monte-Carlo estimator on a wide interval.

Both techniques can be slightly modified for other distributions and can be easily programmed and executed on main frame as well as personal computers. The algorithms and computer programs will be available.

R

THE EFFECT OF SMALL COVARIATE-CRITERION CORRELATIONS ON ANALYSIS-OF-COVARIANCE

M. Rovine, A. von Eye, P. Wood College of Human Development
The Pennsylvania State University, University Park, PA 16802

In uncontrolled studies, those studies in which individuals are not randomly assigned to experimental and control groups but are members of different levels of categorical variables, analysis of variance is most often suggested as the appropriate data analytic tool for assessing group differences on any dependent or criterion variables of interest. When variables may be identified that are related to the criterion variable and may act as plausible, alternative hypotheses analysis of covariance has been suggested. In theory, this analysis may have some effect in "equating" groups according to their scores on the covariate. However, since ANCOVA was designed to increase the precision of randomized experiments, at least two questions arise: 1) Is this technique appropriate in uncontrolled studies? 2) Must the size of the covariate-criterion relationship meet a minimum value? To assess these questions, a simulation was performed to indicate the degree of bias in the analysis of covariance under the condition of low covariate-criterion correlations.

The method used in this study looked at the change in the significance levels of the F-test of the ANOVA by adding a covariate that has a non-zero, but non-significant correlation with the criterion variable. By adjusting for nothing other than sampling fluctuation, an estimate of the degree of bias associated with the inappropriate selection of a covariate was obtained.

To show the degree of bias introduced when controlling for statistically non-significant relationship, a simulation study was run in which a criterion variable was created by generating a random normal variate and assigning a group number (either 1 or 2) to each value of the variate. A constant was then added to the second group to create the group difference. The constant was incremented by .05 until the difference between the groups became statistically significant at the $p<.001$ level. Covariates were then selected by generating a set of random variates and selecting those that had correlations ranging from $r=.01$ to a level just under the $p<.05$ level of significance.

The results of the study showed that by covarying random fluctuation out of a dependent variable, one can artificially decrease the size of F-test denominator. This is tantamount to an arbitrary decision to make the error term of the ANOVA smaller in the absence of any reasonable covariates.

The Effect of Measurement Error in a Machine Learning System

David L. Rumpf
Mieczyslaw M. Kokar

Department of Industrial Engineering and Information Systems
Northeastern University
Boston, MA 02115

ABSTRACT

This paper deals with the problem of reasoning about conceptualizations (sets of relevant parameters) of physical processes. The problem is discussed in the context of the COPER discovery system. COPER conjectures parameters characterizing physical processes and the functional relationships among them. The COPER system utilizes the idea of changing representation base to determine the arguments of invariant functional descriptions. It must handle two types of uncertainty - about relevance of parameters and measurement error. A statistics/probability approach has been used to estimate the effect of measurement error in the COPER system. The partially adequate results of this approach are presented. Alternative approaches to the measurement error problem will be suggested and discussed.

R

Maximum Likelihood Estimation of Discrete Control Processes: Theory and Empirical Applications

John Rust
Department of Economics
University of Wisconsin

Consider the following "identification" or "revealed preference" problem. We observe data generated by agents solving infinite horizon markovian decision problems. At time t each agent observes a vector of state variables (x_t, ϵ_t) and chooses an action i_t from a finite set of alternatives to obtain a reward which depends on (x_t, ϵ_t, i_t) and a vector of parameters θ_1 which are known by the agent but not by us. The state variables evolve according to a markov process with transition density which depends on a vector of parameters (θ_2, θ_3) also known by the agent but not by us. Our data consists of independent realizations $\{i_{tl}, x_{tl}\}$, $t = 1, \ldots, T$, for each agent l, $l=1, \ldots, L$. Our problem is to go "backwards" and use this data to infer the unknown parameter vector $\theta = \{\beta, \theta_1, \theta_2, \theta_3\}$, where $\beta \in (0, 1)$ is the discount factor. This paper derives a nested fixed point maximum likelihood algorithm to estimate the unknown parameters of a subclass of these "discrete control processes". We show that either as T or $L \rightarrow \infty$ the estimated parameter vector θ converges to the true parameter vector with probability 1 and has an asymptotic Gaussian distribution. In order to illustrate the use of the algorithm, we discuss two empirical applications: 1) a model of optimal retirement of bus engines, and 2) a model of optimal retirement of human beings.

Advanced Statistical Computations Improve Image Processing Applications

Bobby Saffari
Generex Corporation

Abstract

Modern computer imaging in conjunction with advanced statistical processing are responsible for significant advances in the areas of medicine and industrial inspection.

Inspections based on the human eye are in many cases tedious, inaccurate, and time consuming. Image processing techniques and computer graphics offer the capability to overcome these set-backs.

The specific area under consideration in this paper is the study of hair density variations over time. Since hair growth and hair loss occur in a non-predictable and random fashion, the human eye is practically incapable of measuring and recording these changes. Statistical processing and computer imaging have been used to facilitate hair density measurement. However, the current techniques have certain shortcomings and flaws.

The purpose of this work is to eliminate the current obstacles and introduce new techniques. These techniques include use of artificial intelligence and local statistical processing such as histogram analysis and Baysian classification criteria. Also methods to eliminate 3-D distortion and enviromental variations are introduced.

Real-Time Classification and Discrimination Among Components of a Mixture Distribution

Douglas A. Samuelson
International Telesystems Corp.

We consider a system in which we collect and analyze, in real time, observations of a statistic with a multimodal (or mixture) distribution. Such distributions arise, for example, in collecting service times when serving multiple classes of customers, each class having a different service-time distribution, at a single service facility. We present new , computationally intensive methods, free of distributional assumptions, to classify current and future observations into one of the underlying classes, and to provide real-time updating of the classification scheme.

Random Graphs

Edward R. Scheinerman
The Johns Hopkins University

An exciting branch of both graph theory and probability is the study of random graphs. In the most popular model of random graphs, the vertices of the graph are fixed and edges are inserted between pairs of vertices at random. Each possible edge is inserted with probability p (or absent with probability $1-p$) and each pair of vertices is considered independently. Because random graphs are easy to generate on a computer, one can perform "experiments" to create and test conjectures about random graphs. We discuss some of our successes and failures in this "experimental" process. Our discussion will include Hamiltonian closure in random graphs and properties of random interval graphs.

S

LINEAR COMBINATIONS OF ESTIMATORS OF THE VARIANCE OF THE SAMPLE MEAN

Bruce Schmeiser

Wheyming Tina Song

School of Industrial Enginering

Purdue University

West Lafayette, IN 47907

(317) 494-5422

(schmeise@gb.ecn.purdue.edu)

We investigate linear combinations of well-known estimators of the variance of the sample mean of strictly stationary time series, including nonoverlapping batch means, overlapping batch means, standardized time series, and spectral-regression estimators. Bias, variance, and mean squared error are examined for various processes, estimator types, and estimator paramters using analytic, numerical, and Monte Carlo methods.

An application of quasi-Newton methods in parametric empirical Bayes calculations

David Scott
Department of Decision Sciences and MIS
Concordia University
Montreal, Quebec H3G 1M8

Abstract

There has been a surge of interest in parametric empirical Bayes methods since Dempster, Laird, and Rubin (1977) showed the applicability of the iterative EM process to hyperparameter estimation. This process is normally computationally intensive, as at each iteration a posterior expectation must be calculated. To reduce computation when the hyperparameter to be estimated is a variance, many researchers (e.g., Wong and Mason, 1985) have used a Gaussian approximation to the posterior distribution at each EM iteration. The estimated posterior mean is then the mode of the posterior, which can be calculated using a Newton-type method for function maximization. In addition, the Gaussian approximation permits the Hessian inverse at the optimum for each iteration to be used to calculate a new estimate of the hyperparameter.

This research investigates the use of a quasi-Newton technique, employing a BFGS update, in the calculation of the posterior mode at each iteration of an EM procedure in an empirical Bayes problem with an unknown prior variance. We maintain only the Cholesky factor of the Hessian, and update this factor using a Householder technique due to Gill, Golub, Murray, and Saunders (1974). Thus we never need to decompose the Hessian, reducing from $o(n^3)$ to $o(n^2)$ the number of arithmetic operations required at each Newton iteration (where n is the number of parameters to be estimated). In addition, the Hessian inverse is readily available through a forward- and back-solution. For empirical Bayes problems involving many parameters, the computational savings can be substantial.

We present computational results from empirical Bayes parameter estimation in a paired-comparison setting.

References

Dempster, A. P., N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, series B, 1977, 39, 1-38.

Gill, P. E., G. H. Golub, W. Murray, and M. A. Saunders. Methods for modifying matrix factorizations. Mathematics of Computation, 1974, 28, 505-535.

Wong, G. Y. and W. M. Mason. The hierarchical logistic regression model for multilevel analysis. Journal of the Americal Statistical Association, 1985, 80, 513-524.

S

Efficient Algorithms for Smoothing Spline Estimation of Functions With or Without Discontinuities

by

Jyh-Jen Horng Shiau
Department of Statistics
University of Missouri - Columbia
Columbia, MO 65211

Abstract

In this paper, we present some efficient algorithms for smoothing spline estimation of an unknown function which is smooth except for some known break points, where discontinuities occur on either the function or its lower order derivatives. For a problem with n observations, these algorithms require $O(n)$ operations for equally spaced knots case and $O(n^2)$ operations for unequally spaced knots case. Similar efficient algorithms are also derived for the ordinary smoothing splines.

Multiply Twisted N-Cubes For Parallel Computing

T.-H. Shiau, Paul Blackwell and Kemal Efe

Department of Computer Science
University of Missouri, Columbia, MO 65211

Abstract: It is known that by twisting one pair of edges of the N dimensional cube, the resulting graph denoted by TQ(N) has diameter N-1 instead of N. In this work, we show that by twisting multiple pairs of edges as well as pairs of buses (a bus is defined as a set of edges with certain common properties), the diameter becomes $\lceil 2N/3 \rceil$. The resulting multiply twisted N-cube, denoted by MTQ(N), preserves most of the desirable topological properties of the ordinary N-cube for parallel computing. A simple routing method is presented which can easily be implemented. Finally we discuss generalizations of MTQ(N) for which the diameters can be made even smaller in the expense of more complicated routing. The smallest diameter which can be achieved by this approach is $\lceil (N+1)/2 \rceil$.

This research is supported in part by AFOSR under Contract AFOSR-86-0124

S

Approximations of the Wilcoxon Test in Small Samples with Lots of Ties

Arthur R. Silverberg

Food & Drug Administration, Rockville, MD

The Wilcoxon-Mann-Whitney Test for two independent samples is frequently used with data having ties. Although, there are computer programs to calculate the exact test, even for small samples computer packages use approximations based upon the normal distribution. Comparisons of the exact and appropriate distributions are found in the literature for a few specific cases. For each of the small sample sizes considered, all distributions of obtaining ties were considered, as well as all permutations of the ordering of the ties. The exact distribution, tabulated value without ties, normal approximations with and without continuity corrections, and Edgeworth expansions with and without continuity corrections, were compared.

Application of Orthogonalization Procedures to Fitting Tree-Structured Models

Cynthia O. Siu
Johns Hopkins University

ABSTRACT

Orthogonalization is an important tool in computations for linear model. In this paper, applications of Givens rotations and Modified Gram-Schmidt orthogonalization to tree-structured regression are discussed. The resulting procedure generalizes CART's piecewise-constant tree model to piecewise linear model. Great versatility is offered by this approach: regression tree models for quantitative and binary data can be handled by one general fitting procedure. In addition, it provides a basis for implementing various linear and tree-structured regression methods under one framework.

An Alternate Methodology for Subject Database Planning

Craig W. Slinkman
Henry D. Crockett
Mark Eakin

University of Texas at Arlington

An important aspect of data administration is strategic data planning. Strategic data planning is the scheme which an enterprise uses to ensure that its information systems function can support the managerial objectives of the enterprise. An important component of strategic data planning is the determination of the subject databases needed. James Martin has suggested a simple ad hoc procedure for performing this analysis. An alternative procedure is suggested using SAS to perform a multivariate statistical technique called correspondence analysis. This technique has the advantages that it has a strong theoretical justification, yields a numerical measure of the strength of the subjective database clustering, and is relatively simple to include in CASE software.

Some Numerical and Graphical Strategies for Implementing Bayesian Methods

Adrian Smith
A. M. Skene
J. E. H. Shaw
J. C. Naylor
S. E. Hills

Summarizing the information in an irregular or multiparameter likelihood in terms of local maxima and curvature may be extremely misleading. However, the routine implementation of integrated likelihood methods requires the development of novel, efficient numerical integration and interpolation strategies, exploiting modern interactive computing and graphics facilities. Progress with the development of such techniques will be reviewed and illustrated.

Variable Selection in Multivariate Multiresponse Permutation Procedures

Eric P. Smith
Department of Statistics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24060

Multiresponse permutation procedures (MRPP) of techniques for analysing data based on the distance betweeen objects. These methods are useful in applications where the number of variables of interest may be large relative to the number of replicates and data may be highly nonnormal. For example, in studies on the bacteria in the mouth there may be as many as 100 possible species, many of them rare.

Besides an overall test of differences between groups, a researcher is usually interested in questions about which variables are important and which groups differ. In this talk some approaches to the problem of variable selection and variable importance are discussed. A stepwise procedure for variable selection is described. Simulation is used to assess and compare the techniques.

S

Gamma Processes, Paired Comparisons, and Ranking

Hal Stern

Harvard University

Models based on gamma random variables for analyzing ranked data are considered. These are natural models for ranking problems in which k objects are ranked according to the waiting time for r events to occur. A sports competition in which the participants are ranked based on the time until a certain number of points are scored is an example of such a problem. For these problems, the probability that k objects are ranked according to a particular permutation can be modeled as the probability that k independent gamma random vavriables with shape parameter r are ranked in that order. Integer values of r describe many common situations. Other values of r are introduced by considering an independent increments Gamma process indexed by r. The value of this process at r can be interpreted as the waiting time until the r^{th} event even when r is not an integer. For each r, a parametric model is developed by considering permutations of the values of k independent Gamma processes with different scale parameters.

The paired comparison problem is a special ranking problem in which only two objects can be compared at a time. The Bradley-Terry and Thurstone-Mosteller paired comparison models are special cases of the Gamma process model, corresponding to r equal one and r tending to infinity. In addition, values of r near zero result in another widely used model. The gamma model provides a unified derivation of these three models and a continuum of new models in between. The gamma models that result from particular choices of r are fit to several paired comparison and ranking data sets.

BAYESIAN ANALYSIS USING MONTE CARLO INTEGRATION -- AN EFFECTIVE METHODOLOGY FOR HANDLING SOME DIFFICULT PROBLEMS IN STATISTICAL ANALYSIS

Leland Stewart

Lockheed Palo Alto Research Laboratory

Both a mathematical and a graphical description of Bayesian analysis using Monte Carlo integration will be presented. The capabilities of this approach will be illustrated by two examples.

In the first example this methodology easily handles rich multiparameter families of univariate distributions; censored, interval and binary data; non-conjugate priors; extrapolation uncertainty; and the computation of posterior distributions for cdf's, hazard rates and densities.

In the second example, this approach allows the statistician to compute the posterior probability for each model in a set of possible models and therefore to retain consideration of several or many models throughout the analysis rather than to restrict attention to just one 'best' model.

Similarities and differences between this methodology and the Bootstrap will be pointed out.

SIMDAT AND SIMEST: DIFFERENCES AND CONVERGENCES

James R. Thompson

Rice University and M.D. Anderson Hospital & Tumor Institute

SIMDAT is an algorithm developed at Rice and the Ballistics Research Laboratory for the empirical simulation of pseudo-data from a data set of high dimensionality. SIMEST is an algorithm developed at Rice and M.D. Anderson Tumor Institute for estimating the parameters of a stochastic process without the generally prohibitive difficulty (in nontrivial cases) of obtaining a closed form for the likelihood. Considerations are given for the use of SIMDAT as a part of the SIMEST algorithm.

S

SIMULATED POWER COMPARISONS OF MRPP RANK TESTS AND SOME STANDARD SCORE TESTS

Derrick S. Tracy and Khushnood A. Khan
Department of Mathematics and Statistics
University of Windsor, Windsor, Ontario, Canada

ABSTRACT

To test the hypothesis of random classification versus classification according to some a priori scheme, Mielke, Berry and Johnson (1976) introduced a test based on multiresponse permutation procedure (MRPP). This test does not require assumptions of normality and homogeneity, and works well for data at ordinal or higher levels. The test statistic is $\delta = \sum_1^g c_i \xi_i$ for g subgroups, c_i is a suitable weight and ξ_i is the average distance for all distinct pairs in the i^{th} subgroup. The distance measure is usually $\Delta_{IJ} = |R(X_I)-R(X_J)|^{\nu}$, where $R(X_I)$ is the rank of X_I in the combined sample. Corresponding to $\nu = 1, 2$, the test statistics δ_1, δ_2 and their simulated power performance have been studied for several underlying populations, e.g., in Tracy and Khan (1987). In this paper, we compare their powers with those of some standard nonparametric tests, for example, normal score and signed score tests. Using extensive simulation, conclusions are drawn for various combinations of sample sizes from several underlying populations.

Mielke, P.W., Berry, K.J. and Johnson, E.S. (1976). Multiresponse permutation procedures for a priori classifications. Comm. Stat. - Theor. Meth. A, 5, 1409-1424.

Tracy, D.S. and Khan, K.A. (1987). MRPP tests in L_1-norm. Comptl. Stat. & Data Anal., 5, 373-380.

Belief Function Computations for Paired Comparisons

David Tritchler
Ontario Cancer Institute and University of Toronto

Gina Lockwood
Ontario Cancer Institute

The theory of belief functions has been used to extend the method of paired comparisons to take into account the varying certainty about the paired choices. These certainties are modelled as belief functions and are incorporated into the analysis of preference structure; the preference model itself is also modelled as a belief function. The conflict between various belief functions is used as a basis for diagnostics describing the choice task.

The computational complexity of the method is high. This paper considers the computational problem. Some shortcuts are obtained using results from the theory of belief functions and graph theory. Monte Carlo methods and the use of symbolic programming are also discussed.

An expert system for prescribing statistical tests of non-parametric and simple parametric designs

Gary W. Tubb
Instructional Computing
University of South Florida
Tampa, FL 33620

An inordinate amount of faculty time is often consumed advising behaviorial science students in the use of appropriate statistical tests. The experimental designs are often straighforward and result in analyses of non-parametric or simple parametric data. This paper describes an expert system written in Turbo PROLOG that prescribes appropriate statistical tests for such simple designs.

The expert system queries the student for example data values of a single subject and the variable name for each data value. Then the system queries for the probable range of the data values. Options for missing data and the transformation of data are provided. The student then identifies the variables to be compared, correlated, tabulated, etc. Based upon this information, the expert system proposes statistical techniques for systematically analyzing the data. The student may query the expert system regarding the logic of employing a specific statistical technique.

Performance of Several One Sample Procedures

David L. Turner
and YuYu Wang

Empirical p-values and powers for the usual t test, the signed rank test, a trimmed t test, a jackknife and a bootstrap procedure were compared using repeated samples of size 30 from normal, double exponential, cauchy, negative exponential and uniform distributions for normal power values ranging from 0.05 through 0.95. The Bootstrap performed as well as the usual t test. The trimmed t, signed rank test and the usual t-test performed about the same. The jackknife performed worst among these tests. The signed rank test did best for the cauchy distribution.

Modeling Parallelism: An Interdisciplinary Approach

Dr. Elizabeth A. Unger, Professor
Kansas State University
Department of Computing and Information Sciences

One can easily conjecture that we humans have imposed sequential solutions onto most problems, such are a better match to our physical architecture, but we propose that there are parallel solutions to many problems and these are a better if they can be matched to our computer architectures. The discovery of problems involving parallelism in many and diverse disciplines which are the subject of current research efforts has been a simple matter, however the development of methods which discover the parallelism possible in solutions to a problem is not a simple matter and is the focus of this research. This paper will describe the model and discuss the current research efforts in terms of academic contributions and the strengths gained through the interdisciplinary group approach to problem solving.

At Kansas State University a group of people from three disciplines in two colleges has been formed to provide a critical mass of researchers and to create broader base of knowledge from which to draw to find an architecture-free model which can be used to express, in a natural way, the potential concurrency in problem solutions. A partially defined model based upon a conditioned dataflow which incorporates the concepts of control flow based on dataflow, of the description of an action at any level of detail with subsequent further refinement if desired, of repetition based upon partitions of data aggregates, of single assignment of values to uniquely identify each incarnation of data objects, and of partial computation, i.e., computation which can proceed until a needed unavailable datum is encounter has been developed. The group has four major foci to their work, 1) continuing development of the theoretical foundation of the model, led by the computer scientists, 2) use of the model to discover paradigm parallelism models for particular problems at the small and the large granularity levels of detail, led by the statistician and engineers, 3) the development of methods of determining the best fit of the disovered parallelism to existing architectures, led by the statistician and engineers, 4) the continued implementation of a prototype on a distributed network of processors, led by the computer scientists. All members have contributed to all phases.

The current status of our work includes a model which has been shown to contain a core of statements which always describe determinate problem solutions for atomic data types. A prototype of the model is operating, albeit a bit inefficiently at the present time, on a network of loosely coupled processors. The prototype is being used to study problem solutions where the granularity of the parallelism is small. On going research work involves providing the theoretical basis for temporally partitioned data aggregates, the inclusion in the prototype of partial computation, and limited data structures and the development of models of existing architectures using the model for the current multiprocessor architectures.

Some Statistical Problems in Meteorology

Grace Wahba
Depatment of Statistics
University of Wisconsin

I will discuss some statistical problems that arise in merging data from various sources to provide estimates of the current state of the atmosphere, for the purpose of providing initial conditions for numerical weather prediction. Some interesting theoretical statistical questions arise. Of course the practical and theoretical questions only sometimes come together - meteorological data can be very messy and have error structure that can be hard to model. Other challenges concern the blending of physical and prior statistical information, the numerical problems inherent in the simultaneous analysis of extremely large data sets, the detection of unreliable forecast,(etc.).

Encoding and Processing of Chinese Language -- A Statistical Structural Approach

Chaiho C. Wang
U.S. Department of Justice and The George Washington University
Washington D. C. 20001

ABSTRACT

Efficient encoding of an ideographic based language, such as Chinese, depends on two key factors: statistical structure of the language and pattern recognition technology. Statistical analysis and computer technology must evolve hand-in-hand. This paper proposes procedures that incorporate user friendly input schemes with low redundancy internal coding methods for computer storage. Attempts are made to integrate the traditionally divided phonic and graphical methods. Special attention is paid to minimizing human effort in the total word processing process.

ON COVARIANCES OF MARGINALLY ADJUSTED DATA

April, 1988

James S. Weber, Asst. Prof., Dept. of Mgmt, Roosevelt University, Chica IL* (*- Preferred mailing address: PO Box 603, Gurnee, IL 60031-0603.

1980 AMS Subject Classification: Primary 62-07, 62-04
Secondary 62A10, 62D05, 62H17, 62N99, 62P20, 62P25

Key Words and Phrases: iterated proportional fitting algorithm, (IPFA), contingency table, interaction matrix, diagonally equivalent matrices.

ABSTRACT

The adjustment of contingency tables to have prescribed row and column sums occurs frequently in applications. (Eg. adjustment of a cross classified sample; trip distribution & migration modeling; certain budget allocation techniques; etc.)

If there is uncertainty and a covariance structure associated with the marginal sums and with the interaction matrix, then it may be desirable to know how this variability propagates to the scaled interaction matrix.

We describe this propagation with approximate covariances obtained from derivatives of the scaled matrix in a linear function of the covariances of the independent variables.

A number of complications make this effort interesting. 1. The scaled interaction matrix is implicitly defined function of the initial interaction matrix or the row and column sums. The derivatives require either an inverse of a singular matrix or an iterative procedure. Here we chose an iterative procedure (and describe the convergence carefully.) 2. There is a functional dependence among the row and column constraints. Obviously this is related to the singular matrix mentioned in #1, but in applications this dependence must be specified behaviorally rather than mathematically.

The contributions of the proposed paper are: 1. We explain an iterative procedure for computing the derivatives of the Iterated Proportional Fitting Algorithm ("IPFA") for interaction matrices with specified marginal sums which properly reflects the functional dependence between row sums and column sums; 2. We clarify that there IS a dependence of the covariances of the marginally adjusted data upon the way in which the dependence of the row and column sums is specified so that the sum of the row sums equals the sum of the column sums; 3. We discuss several ways of insuring row and column sum consistency; 4. We provide approximate expressions in a factored form showing in detail the sensitivities to the variability of each of the independent variables. (Simulations do not give this level detail.)

W

Bayesian Diagnostics for Almost Any Model

Robert E. Weiss
University of Minnesota

When calculating a Bayesian posterior mean using a numerical method such as Monte Carlo or Quadrature, it is very easy to also compute influence and outlier case statistics for each data point at small extra cost. Most of the Bayesian diagnostics currently in the literature are functions of the predictive distribution of the next data point. This leads to the predictive plot, a graph of the predictive distribution of the next observation as a covariate changes. Predictive plots can be used for model checking in addition to the obvious use as a prognostication.

Variants of Tierney-Kadane

G. Weiss & H. A. Howlader

University of Winnipeg, Winnipeg, Manitoba

Abstract

Bayes estimation of the reliability function of the logistic distribution under a log-odds squared error loss with a non-informative prior is considered by using the approximation method of Tierney & Kadane (1986). *Direct application of the procedure does not yield correct results and so some variations of the procedure are considered.*

Session: Inference and Expert Systems

COSTAR : An Environment for Computer-Guided Data Analysis

by
David A. Whitney
Ilya Schiller

The Analytic Sciences Corporation
55 Walkers Brook Drive
Reading, MA 01867 (617) 942-2000

This paper describes work in progress on the development and implementation of COSTAR, a tool for COordinated STatistical Analysis and Reasoning.COSTAR illustrates the integration of high-end symbolic / numerical hardware and software environments. One objective of this work is to use modern "off-the shelf" statistical and expert-systems programming tools that allow the developer to focus more on the content of the system, and less on implementation details. Symbolic processing is implemented in KEE and Common Lisp on a Symbolics workstation, with numerical processing performed on a mini-supercomputer, the Alliant FX/8 running IMSL and Fortran. The knowledge base uses frames to represent a hierarchy of data objects and directs the development and application of rules through the use of rule classes. The system implements such a rule-based inferencing system for ARIMA time series modeling.

COSTAR is designed to be a tool both for solving statistical problems, and for studying strategies for solving data analysis problems. In this regard, it owes an intellectual heritage to both REX and DINDE. The system development perspective here is primarily that of a statistician, not of an AI scientist. The system is designed for a fairly sophisticated user who can be expected to contribute to parts of the analysis -- an interactive, graphical, two-way user interface is an important part of the system. The system leverages the user's ability and increase efficiency by executing routine analysis, presenting the user with options when decisions are not clear-cut, and asking for user-input if new situations are encountered. The system provides for trace or logging facilities to keep track of analysis sessions. These traces are used to help refine data-dependent statistical strategies, and to support the refinement, formalization, and "learning" of rules in the knowledge base. Such traces also play an important role in the validation of the inferencing schemes in the system. It is designed as a system which will start with basic expertise in a data analysis method, but that is also able to acquire specific applications expertise as analysis sessions are recorded and reviewed.

This paper describes the a itecture of the prototype COSTAR system and the ARIMA modeling knowledge base implemented. System validation procedures are discussed, along with the trace facility for analysis cataloging and rule refinement. Plans for study of more sophisticated, more automatic rule refinement schemes are also discussed.

Bayes Estimation of Cerebral Metabolic Rate of Glucose in Stroke Patients

P David Wilson, SC Huang, RA Hawkins

Local cerebral metabolic rate of glucose (LCMRG) in a local region [redacted] human brain can be calculated as a nonlinear function of the ra[redacted] constants in a 3-compartment model. The model describes the fate of deoxyglucose (DG) in the region following injection into a peripheral vein. The compartments are: (1) DG in plasma, (2) free DG in brain tissue, and (3) phosphorylated DG in brain tissue. If the injectected DG is labeled with Fluorine-18, a positron emitter, a positron emission tomography (PET) scanner can record the relative concentration of the F-18 label in the region. To a close approximation the contribution of compartment (1) to the PET data can be ignored, and the PET data can be said to represent a noisy version of the combined contributions from compartments (2) and (3). From a linear systems viewpoint, the F-18 concentration versus time function in the combined compartments (2) and (3) can be viewed as the output function of a system in which the impulse response is a biexponential time function witn coefficients (called macroparameters) which are nonlinear functions of the rate constants. The input to the system is the concentration versus time function of F-18 in compartment (1), and this can be observed in a peripheral vessel. The output function is the convolution of the impulse response and the input function. If the input and output functions are observed repeatedly over a 2.5 to 3 hour period after injection, nonlinear regression methods can be used to estimate the macroparameter coefficients of the biexponontial impulse response, and from these the LCMRG can be estimated. However, the long scanning period required is seen as unacceptable for routine clinical studies because the patient is required to lie in the scanner without moving his head for the entire period and because of demand for scanner time. Thus a procedure is desired which will estimate LCMRG from a PET observation at a single time and the input function observed up to that time. Several such "single scan" methods are currently in clinical use. These methods use the values of estimates of the population mean rate constants (but are not Bayes procedures). The rate constants are different in normal and stroke regions of the brain, and preliminary perfusion scans and transmission computed tomography scans would be required to delineate the stroke region of the brain. But LCMRG estimation procedures are desired to be independent of such preliminary scans, and the existing single scan methods make large systematic errors in stroke tissue when using mean rate constant values for normal tissue. We developed a Bayes procedure for use with a single scan. Empirical prior mean vectors and covariance matrices are available for the macroparameters for both normal and stroke tissure separately. Empirical prior results are also available for the error variance of the PET observations. For each tissue type, we assumed that the macroparameters are Gaussian distributed among individuals in the population and that the reciprocal error variances are gamma distributed. The Bayes procedure computes the posterior distribution of the macroparameters twice, once using the prior density for each tissue type, and selects the macroparameter estimates associated with the highest posterior density. We conducted computer simulation studies to display the behavior of the Bayes procedure for stroke tissue and to compare it with the other single scan methods. Mean and root-mean-square percent errors are given for a range of true LCMRG values in stroke tissue. The Bayes procedure is seen to be superior to the other methods.

NETWORKS TO SUPPORT SCIENCE

Stephen Wolff
National Science Foundation
1800 G Street, N. W.
Washington, D. C. 20550
(Electronic Mail: steve@cerberus.DNCRI.NSF.GOV

ABSTRACT

More than 150 academic, industrial, and government research campuses are now attached to NSF-sponsored, mid-level computer networks and interconnected by the transcontinental NSFNET Backbone Network. The connection of multiple supercomputers to the Backbone has extended high performance computing to the largest constituency ever; in particular, more statisticians than ever before can be Practicing - as well as Thinking - the Unthinkable.

Of equal, and in the long run even greater, importance is that the transparent connection of the NSFNET family of networks and the ARPANET (achieved by joint adoption of an open protocol set) has achieved a critical level of scientist-to-scientist connectivity. Just as highways and railroads enabled the ready assemblage and interaction of raw material, capital, and labor to fuel the Industrial Revolution, so the emerging National Research Internet is enabling intellectual concentrations of unprecedented scale and agility, and a new epoch of the Information Revolution based on Collaboration Technology is underway.

All-Subsets Regression on a Hypercube Multiprocessor

Peter C. Wollan
Department of Mathematics
Michigan Technological University
Houghton, MI 49931

All-subsets regression (that is, computing linear regressions for all subsets of k predictors) is an inherently parallel problem, suitable for exploring the use of hypercube multiprocessors in statistical computation. The algorithm described here uses the sweep operator for introducing or removing variables; the load is apportioned among processors in a nearly optimal way, based on the Gray code embedding of a hypercube into a torus. The algorithm is implemented in FORTRAN on an Intel iPSC d4. The program's general behavior suggests that while hypercube multiprocessors are potentially valuable for data analysis, their use will require development of new methods.

Enclosure 1
9470:R3012

W

An Iterative Bayes Method for Classifying Multivariate Observations

Duane E. Wolting
Aerojet TechSystems Company

ABSTRACT

A method is presented for classifying multivariate observations. The method uses a Bayes decision rule, which is initially determined from a sample of training observations. Subsequent observations classified with this decision rule are used to adjust the rule in a nonsupervised fashion. These same observations are then reclassified using the adjusted decision rule. The process is repeated until convergence is attained.

The behavior of this algorithm is examined in a series of computer simulation studies. The effects of interclass separation, training sample size, number of classes and dimensionality are considered. The results suggest that under certain conditions this method reduces the misclassification rate by as much as 30%. Although computationally intensive, the algorithm appears to converge in relatively few iterations. Applications to pattern recognition are discussed.

KEYWORDS: Bayesian estimation, classification, computationally intensive methods, decision-theoretic recognition, iterative procedures, nonsupervised learning, pattern recognition.

On the Convergence of Variable Bandwidth Kernel Estomators of a Density Function

Ting Yang
University of Cincinnati

We consider here the Rosenblatt-Parzen kernel estimators of an unknown density function, but this time with a variable (local) bandwidth. The consistency is studied for variable bandwidth kernel estimators. We also have simulated and shown that *in terms of integrated mean squared error* (for any sample size), the kernel estim.tors with local bandwidth cnoice are better than the ordinary kernel estimators with global bandwidth if optimal bandwidths are used.

A COMPARISON OF SEVERAL METHODS FOR GENERATING EXPONENTIAL POWER VARIATES

Dean M. Young
Baylor University

Danny W. Turner
Baylor University

John W. Seaman
University of Southwestern Louisiana

ABSTRACT

This paper compares several alternative algorithms for generating observations from an exponential power distribution with parameter r, $1 < r < 2$. The algorithms include squeeze methods, a ratio-of-uniforms method, and an almost-exact inversion method. A comparison of marginal execution times is made among the variaous methods mentioned above and the generalized acceptance/rejection method proposed by Tadikamalla (1982).

Participant Index

(includes speakers, co-authors, session chairs and program committee members)

Appendix B
List of Paid Attendees

20th INTERFACE SYMPOSIUM

PAID REGISTRANTS

Ahsanullah M.
Department of D.S. and Computation
Rider College
2083 Lawrenceville Road
Lawrenceville NJ 08648-3099
609-885-5530

Ait-Kheddache Amar
Electrical and Computer Engineering Dept
North Carolina State University
Box 7911
Raleigh NC 27695-7911

Allen Joseph
Rm 4168
USDA, NASS, SRB, R&AD
14th & Independence Ave., SW
Washington DC 20250
(202) 447-3778

Alm Barney
Numerical Algorithms Group Inc.
1101 31st Street, Suite 100
Downers Grove IL 60515
(312) 971-2337

Almond Russell
Department of Statistics
Harvard University
1 Oxford Street, Science Center 609
Cambridge MA 02138
almond@hustat.harvard.edu
(617) 495-4888

Altman Naomi
Cornell University
Ithaca NY

Anandalingam Professor G.
Department of Systems
University of Pennsylvania
Philadelphia PA 19104-6315
anand@eniac.seas
(215) 898-8790

Anneberg L. M.
Wayne State University
3123 Roosevelt
Dearborn MI 48124
(313) 565-2763

Atilzan Taskin
Room 30J-014
AT&T Bell Labs
600 Mountain Avenue
Murray Hill NJ 07974-2070
(201) 582-2824

Atkinson E. Neely
Department of Biomathematics, Box 237
University of Texas System Cancer Center
1515 Holcombe Blvd.
Houston TX 77030
an123651@unhvml.bitnet
(713) 792-2619

Aubuchon Jay
Minitab, Inc.
3081 Enterprise Drive
State College PA 16801
ATU@PSUVM.BITNET
(814) 238-3280

Ball Celesta G.
5225 Bradfield Drive
Burke VA 22015
(703) 425-5846

Banchoff Thomas
Department of Mathematics
Brown University
Providence RI 02912
(401)863-1129

Banfield Jeff
Department of Mathematical Sciences
Montana State University
Bozeman MT 59717-0001
UMSFJBAN@MTSUNIX1
(406) 994-5367

Baran R.H.
Intellimetrics, Inc.
4508 Cheltenham Drive
Bethesda MD 20814

Barlow Richard
Operations Research Center
University of California
Etcheverry Hall
Berkeley CA 94720
airforce@violet.Berkeley.EDU
(415)642-4993

Barnhill Bruce M.
1014 Farragut Street
Pittsburgh PA 15206

Barr Jewel T.
South Agriculture Building
USDA-NASS
14th & Independence Avenue, SW
Washington DC 20250
(202) 447 7611

Barron Andrew R.
Department of Stat. & Electrical/Com Eng
University of Illinois
825 S. Wright Street
Champaign IL 61820
(217) 333-6216

Barron Austin
Department of Math/Stat
The American University
4400 Massachusetts Avenue
Washington DC 20016
(202) 885-3120

Barron Roger L.
Barron Associates, Inc.
Route 1, Box 159
Stanardsville VA 22973-9511
(804) 985-4400

Barthel Michael
727 Monroe St., #102
Rockville MD 20850
(202) 673-6521

Basinski Antoni
321 Rose Park Drive
Toronto Ontario M4T 1R8
CANADA
(416) 488-2712

Basu Asit
Department of Statistics
University of Missouri-Columbia
328 Math Sciences
Columbia MO 65211
(314) 882-8283

Baxter Ron
SCIRO Division of Math & Stat
PO Box 218
Lindfield NSW 2070
AUSTRALIA
ronb@natmlab.oz.au
61-2-676059

Becker Richard
Statistical & Data Analysis Research
AT&T Bell Labs
600 Mountain Avenue, Rm 2C259
Murray Hill NJ 07974
rab@research.att.com
(201) 582-5512

Bedewi Gabrielle
1023 Brice Road
Rockville Md 20852
(301) 424-1481

Bell James M.
M-233
Milliken & Company
P.O. Box 1926
Spartanburg SC 29304
(803) 573-1627

Benedict Jeffrey P.
8614 Spruce Run Court
Ellicott City MD 21043
(301) 461-7586

Berk Kenneth
Department of Mathematics
Illinois State University
313 Stevenson Hall
Normal IL 61761
(309) 438-8781

Billard Lynne
Department of Statistics
University of Georgia
Athens GA 30602
(404) 542-5232

Birkett Tom
USDA S. Building
Department of Agriculture
14th & Independence, S.W. Rm. 5819
Washington DC 20250
(202) 447-5359

Boggs Paul
Statistical Engineering Division
National Bureau of Standards
Room A37/Admin Building
Gaithersburg MD 20899
Boggs@cam-vax.arpa
(301) 975-3816

Bolorforoush Masood
Center for Computational Statistics
George Mason University
4400 University Drive
Fairfax VA 22030
(703) 764-6181

Bolstein Richard
Center for Computational Statistics
George Mason University
242 Science Technology Building
Fairfax VA 22030
(703) 323-2730

Boudreau Robert M.
1043 Finchley Place
Richmond VA 23225
(804) 257-1301

Box George E. P.
Department of Statistics
University of Wisconsin
1210 West Dayton
Madison WI 53706
(608) 262-2937

Brandenburg Joseph
Intel Scientific Computers
15201 NW Greenbriar Parkway
Beaverton OR 97006
joeb@isc.intel.com
(503) 629-7701

Bravo Maria Soledad
Bureau of the Census
SRD
Washington DC 20233
(301) 763-3950

Brooks Daniel G.
Department of Dec. & Inf Systems
Arizona State University
Tempe AZ 85287-4206
ATDGB@ASUACAD
(602) 965-6350

Bross Neal
Bureau of the Census
Department of Commerce
Rm 3217 FOB #4
Washington DC 20233
(301) 763-4466

Brown Barry W.
Department of Biomathematics, Box 237
University of Texas System Cancer Center
1515 Holcombe Blvd.
Houston TX 77030
anl2bwbl@uthvml.bitnet
(713) 792-2614

Brown C. Hendricks
Department of Biostatistics
Johns Hopkins University
Baltimore MD 21218
(301) 855-2420

Brown Morton B.
Department of Biostatistics
The University of Michigan
109 South Observatory
Ann Arbor MI 48109-202(
(313) 936-0992

Brownstone David
School of Social Sciences
University of California
Irvine CA 92717
(714) 856-6231

Bruce Andrew
395 Cherry Lane, R.R.1
Mendham NJ 07945
deborah@flash.bellcore.com
(201) 879-4319

Bugl Paul
University of Hartford
West Hartford CT 06117
(203) 653-6839

Burg John
Entropic Processing, Inc.
1011 North Foothill Blvd.
Cupertino CA 95014-5649
(408) 973-9800

Burn David A.
BBN Software Products Corporation
10 Fawcett Street
Cambridge MA 02238
(617) 873-4249

Campbell Gregory
LSM, DCRT
National Institute of Health
Bldg. 12A, Rm 3045
Bethesda MD 20892
GGC@NIHCU.BITNET
(301) 496-6037

Carlin Brad
Department of Statistics
University of Connecticut
Box U-120
Storrs CT 06268
(203) 486-4312

Carney John B.
USDA, NASS, RAD, SRB. Rm. 4801
14th & Independence Avenue, SW
Washington DC 20250
(202) 475-3675

Carr Daniel B.
Bettelle Pacific Northwest Laboratories
P.O. Box 999,K6-45
Richland WA 99352
(509) 376-3344

Cassells J. Sally
BBN Software Products
10 Fawcett Street
Cambridge MA 02238
(617) 873-8164

Chaloner Kathryn M.
Department of Applied Statistics
University of Minnesota
352 Classroom-Office Bldg.
St. Paul MN 55108
(612) 625-8722

Chamas N. H.
Wayne State University
7816 Bingham
Dearborn MI 48126
(313) 582-8218

Chambers John M.
AT&T Bell Laboratories
600 Mountain Avenue, Rm. 2C-282
Murray Hill NJ 07974-2070
(201) 582-2681

Chandra Mahesh
Department of Bus. Comp. Inf. Sys.
Hofstra University
201 Computer Center
Hempstead NY 11550
(516) 560-5717

Chandrasekar V.
Department of Electrical Engineering
Colorado State University
Fort Collins CO 80521
ERGU764@CSUGREEN.BITNET
(303) 491-6567

Chari Ravi
Department of Mathematics
Tufts University
Medford MA 02155
rchari@amber.cc.tufts.edu
(617) 628-5000

Chen Zehua
129 Mansfield Street
New Haven CT 06511
(203) 776-0148

Chernoff Herman
Department of Statistics
Harvard University
Cambridge MA 02138
(617) 495-5462

Chin Daniel C.
Applied Physics Laboratory
The Johns Hopkins University
Johns Hopkins Road
Laurel MD 20707-6099
(301) 953-7100

Choo Chang Y.
Electrical Engineering
Worcester Polytechnic Institute
Worcester MA 01609
cychoo@wpi.bitnet
(617) 793-5000

Chow Mo Suk
Math Department
Northeastern University
Lake Hall
Boston MA 02184
(617) 843-0651

Cibulskis John M.
SPSS, Inc.
444 N. Michigan Avenue
Chicago IL 60611
(312) 329-2400

Ciuffini Mary Ann
Room 4169 So. Bldg
USDA, NASS, SRB RAD
14th and Independence Ave., SE
Washington DC 20250
(202) 447-6751

Clarkson Douglas B.
2500 ParkWest Tower One
IMSL
2500 CityWest Boulevard
Houston TX 77042-3020
(713) 782-6069

Cofer John
200 Stokely Management Center
UTCC
Knoxville TN 37996
Cofer@UTKVXI
(615) 974-6831

Cohen Edgar A.
White Oak
Naval Surface Warfare Center
New Hampshire Avenue
Silver Spring MD 20903-5000
(202) 294-2281

Coleman Chuck
2300 Lee Highway, #102
Arlington VA 22201
(703) 527-7699

Conlon Dr. Michael
Department of Statistics
University of Florida
Box J-212, J. Hillis Miller Health Centr
Gainesville FL 32610
(904) 392-2820

Cooper Murray M.
Biomathematics Group 7293-32-2
The Upjohn Company
301 Henrietta Street
Kalamazoo MI 49001-0199

Cox Dennis
Department of Statistics
University of Illinois
725 So. Wright Street
Champaign IL 61820
(217) 333-2972

Crane Giles
Office of Research
NJ State Department of Health
John Fitch Plaza, CN360
Princeton NJ 08625
(609) 292-4315

Creecy Rob
SRD 3215-4
Census Bureau
Washington DC 20233
(301) 763-4466

Crockett Henry D.
Department of Info Syst. and Management
The University of Texas at Arlington
Box 19437
Arlington TX 76019
B718MED@UPARLVM
(817) 273-3502

Cybenko George
60 Nowell Road
Melrose MA 02176
cybenko@cs.tufts.edu
(617) 381-3214

Dadashzadeh Mohammad
Department of Management Science
University of Detroit
4001 W. McNichols Road
Detroit MI 48221
(313) 927-1237

Davis Linda
11539 Mamie Lane
Fairfax Station VA 22039
(703) 239-0892

De Los Reyes Josephina P.
Department of Mathematical Sciences
University of Akron
Akron OH 44325
(216) 375-7193

Denby Lorraine
AT&T Bell Laboratories
600 Mountain Avenue, Room 2C273
Murray Hill NJ 07922
research!alice!ed
(201) 582-3292

Deng Lih-Yuan
Department of Mathematical Sciences
Memphis State University
Memphis TN 38152
(901) 454-3141

Devlin Thomas F.
Math & Computer Science Department
Montclair State College
Upper Montclair NJ 07043
(201) 893-7244

Dixon Dennis
National Cancer Institute, NIH
Landow Building, Room 4B06
Bethesda MD 20892
(301) 496-4836

Do Kim-Anh
Department of Prevention Research & Bio
Bowman Gray School of Medicine
211-7 Dalewood Drive
Winston-Salem NC 27104
KIM@WFU.EDU
(919) 748-6015

Dombroski Marla
N.I.H.
Federal Bldg, Room 119
Bethesda MD 20892
mpb@nihcu
(301) 496-9725

Don Dr. Allen
77 Mill Spring Road
Manhasset NY 11030
(516) 627-0098

Donnell Deborah
Bell Core Communications Research
435 South Street
Morristown NJ 07960-1961
(201) 829-4319

Dreher Douglas M.
Hughes Aircraft Company
8000 E. Maplewood Avenue, MS CHSB
Englewood CO 80111

Drummey Kevin Ward
National Security Agency
Attn: R51
Fort Meade MD 20755
(301) 859-6461

DuMouchel William
BBN Software Products Corporation
10 Fawcett Street
Cambridge MA 02238
dumouche@SPCA.BBN.COM
(617) 873-8119

DuPont Pierre
BBN Advanced Computers
800 Westpark Drive
McLeane VA 22101
pdupont@bbn.com
(703) 848-4874

Dunn Jeffrey
Code 2842
Naval Research Laboratory
4555 Overlook Avenue, NE
Washington DC 20375-5000
Dunn@NRL.arpa
(202) 767-2884

Duong Dr. Quang Phuc
Bell Canada
1050 Beaver Hall, Room 205
Montreal Quebec H2Z 1S4
CANADA
(514) 870-4923

Durst Mark
Lawrence Livermore National Lab.
L-307, P.O. Box 808
Livermore CA 94550
Durst@clara.llnl.gov
(415) 422-4272

Dykstra Richard L.
Department of Statistics
University of Iowa
Iowa City IA 52242
(319) 335-0823

Eakin Mark
Department of Info Sys & Mgmt Sciences
The University of Texas at Arlington
Box 19437
Arlington TX 76019
B718MEE@UTARLVM1
(817) 273-3502

Eddy William
Department of Statistics
Carnegie-Mellon University
Pittsburgh PA 15213
bill@andrew.cmu.edu
(412) 268-2725

Edwards Don
Department of Statistics
University of South Carolina
Columbia SC 29208
(803) 777-5073

Edwards Lynne K.
323 Burton Hall
University of Minnesota
178 Pillsbury Drive, SE
Minneapolis MN 55455-0211
(612) 624-8381

Efron Bradley
Department of Statistics
Stanford University
Sequoia Hall
Stanford CA 94305
(415) 497-2206

Elder, IV John F.
Barron Associates
Rt. 1 Box 159
Stanardsville NJ 22973-9511
(804) 985-4400

Elfenbein Lowell
TRW
One Federal Systems Park Drive
Fairfax VA 22033
(301) 776-4920

Engelman Laszlo
BMDP Statistical Software, Inc.
1440 Sepulveda Blvd
Los Angeles CA 90025
(213) 479-7799

Ercil Aytul
Math Department
General Motors
General Motors Labs
Warren MI 48098

Ferguson Dania P.
Room 5862-S
USDA-NASS-DMD
1400 Independence Avenue
Washington DC 20250
(202) 447-6690

Feuerverger Andrey
Department of Statistics
University of Toronto
Toronto Ontario M5S 1A1
CANADA
ANDREY@UTSTAT/UUEP
(416) 978-3452

Feygin Leonid
SPSS, Inc.
444 N. Michigan Avenue
Chicago IL 60611
(312) 329-3544

Filliben James
Statistical Engineering Division
National Bureau of Standards
Rm A337/Admin Bldg
Gaithersburg MD 20899
(301) 975-2855

Fischer Martin J.
DCEC, Code R700
1860 Wiehle Avenue
Reston VA 22090-5500
fischer@EDN-VAX.ARPA
(703) 437-2196

Folk Roni
National Cancer Institute
7910 Weedmont Avenue
Bethesda MD 20892
(301) 496-6425

Franklin Dr. Leroy A.
Department of Systems and Decision Sci.
Indiana State University
Terre Haute IN 47809
(812) 237-2092

Friedman Jerome H.
Stanford Linear Accelerator Center
P.O. Box 4349, Bin 88
Stanford CA 94309
JHF@SLACVM
(415) 926-2256

Furnas George
Bell Communications Research
455 South Street
Morristown NJ 07960
gwf@bellcore.com
(201) 829-4289

Garbo Martin J.
Mail Station CHSB
Hughes Aircraft Company
8000 E. Maplewood Avenue
Englewood CA 80111
(303) 341-3602

Geissler Dr. Paul H.
Patuxent Wildlife Research Center
Route 197
Laurel MD 20708

Gentle James E.
IMSL, Inc.
2500 City West Blvd. Suite 2500
Houston TX 77036
(713) 782-6060

Flournoy Nancy
1829 E Capitol St. SE
Washington DC 20003-1711
nflournoy@note.nsf.gov
(202) 357-3693

Fong Duncan K. H.
The Pennsylvania State University
340 Beam Business Admin Building
University Park PA 16802
I2V@PSUVM
(814) 863-3541

Friedman Herman
IBM Technical Education
500 Columbus Avenue
Thornwood NM 10594
(914) 742-5647

Fritsvold John D.
Washington Navy Yard
Naval Weapons Engineering Support Act
ESA-31E, Bldg. 220-2
Washington DC 20374-2203
(202) 433-3116

Gantz Professor Donald T.
Center for Computational Statistics
George Mason University
ST 242
Fairfax VA 22030
(703) 323-2711

Gardenier Dr. T. K.
115 St. Andrews Drive
Vienna VA 22180

Gelfand Alan E.
Department of Statistics
The University of Connecticut
196 Auditorium Road, U-120, MSB 428
Storrs CT 06268
(203)-486-3416

Gerling Thomas
Applied Physics Laboratory
Johns Hopkins Road
Laurel MD 20707

Geweke John
Department of Economics
Duke University
Durham NC 27706
(919) 684-2152

Gilette Robert E.
Department of the Treasury
7217 15th Avenue
Tacoma Park MD 20912
(301) 422-9227

Gilroy Edward J.
U.S. Geological Survey
410 National Center
Reston VA 22029
(703) 648-5714

Gladstein David
ICAD, Inc.
1000 Massachusetts Avenue
Cambridge MA 02138
(617) 868-2800

Glaz Joseph
Department of Statistics
University of Connecticut
U-120
Storrs CT 06268
(203)486-3413

Goel Prem K.
Department of Statistics
Ohio State University
1958 Neil Avenue
Columbus OH 43210-1247
goel@osupyr.mast.ohiost. edu
(614) 292-8110

Goicoechea Professor Ambrose
Department of Systems Engineering
George Mason University
4400 University Drive
Fairfax VA 22030
(703) 323-3530

Goldman Robert
Simmons College
300 Fenway
Boston MA 02115
(617) 738-2167

Goldstein Larry
Dept of Math, DRB 306 VSC
University of Southern California
University Park - MC 1113
Los Angeles CA 90089-1113
(213) 743-2567

Gray Mary W.
Department of Math Stat
American University
Washington DC 20016
(202) 885-3170

Green Ivan
Statistics Canada

Greene Michael
Math/Stat/Computer Science Department
The American University
4400 Massachusetts Avenue, N. W.
Washington DC 20016
(202) 885-3120

Gu Chong
Department of Statistics
Yale University
Box 2179 Yale Station
New Haven CT 06520
gu-chong@yale.arpa

Gupta Shanti S.
Department of Statistics
Purdue University
W. Lafayette IN 47907
tej@l.cc.purdue.edu
(317) 494-6031

Habib Muhammad
Department of Biostatistics
University of North Carolina
Chapel Hill NC 27514
(919) 966-4887

Hagwood Charles
Statistical Engineering Div
National Bureau of Standards
714-101/A334
Gaithersburg MD 20899
(301) 975-2846

Hahn Dr. Gerald J
GE CRD
P. O. Box 8
Schenectady NY 12301
(518) 387-5558

Hallahan Charles
USDA/ERS
1301 New York Avenue, ND, Rm. 250
Washington DC 20005
(202) 786-1507

Hardle Wolfgang
Department of Economics
University of Bonn
Bonn D-5300
WEST GERMANY
OR624@DBNUOR1.BITNET

Harner James
West Virginia University
215A Morgan Drive
Morgantown WV
(304) 599-9583

Harte James M.
560 N St. SW, N512
Washington DC 20024

Heard Edward
Ciba-Geigy Corporation
556 Morris Avenue
Summit NJ 07901
(201) 277-7085

Hearne Leonard
1829 E. Capital Street, SE
Washington DC 20003
(202) 546-7551

Heckler Charles
Eastman Kodak Company
MQA, B56, Fl 4, KP
Rochester NY 14650
(716) 533-1797

Heiberger Richard M.
Department of Statistics
Temple University
Speakman Hall
Philadelphia PA 19122
(215) 787-6879

Henstridge Dr. John
38 Wellington Street
Oxford OX2 6BB
ENGLAND
0865-510724

Herbert John H.
2929 Rosemary Lane
Falls Church VA 22042
(703) 532-4544

Herzog Thomas N.
U.S.Department of Housing/Urban Dev.
Washington DC 20410-8000

Hietala Paula
Department of Mathematical Sciences
University of Tampere
P.O. Box 607
Tampere SF-33101
FINLAND
358-31-156111

Hill Dr. Mary Ann
BMDP Statistical Software, Inc.
1440 Sepulveda Blvd. Suite 316
Los Angeles CA 90025
(213) 479-7799

Hill Joe R.
EDS Research
2155 Louisiana Blvd, NE Suite 9100
Albuquerque NM 87110
(505) 883-6931

Holland Susan
Suite 500
Mathematica Policy Research Inc.
600 Maryland Avenue, SW
Washington DC 22024
(202) 484-9220

Hormel Phillip
Bldg. D-2078
CIBA/GEIGY Corporation
556 Morris Avenue
Summit NJ 07901
(201) 277-7436

Hsieh John J.
Department of Preventive Medicine & Bio
University of Toronto
McMurrish Building, 4th floor
Toronto Ontario M5S 1A8
CANADA
(416) 978-5203

Hudak Gregory B.
Lincoln Center, Suite 106
Scientific Computing Associates
4513 Lincoln Avenue
Leslie IL 60532
(312) 960-1698

Hurley Catherine
Department of Statistics & Act. Sci.
University of Waterloo
Waterloo Ont N2L 3G1
CANADA
(519) 888-4505

Ihaka Ross
Statistics Center
MIT
1 Amherst Street
Cambridge MA 02139
IHAKA@DOLPHIN.MIT.EDU
(617) 253-8404

Jayachandran Toke
Naval Postgraduate School
Monterey CA 93943
(408) 646-2600

Johnson Laura
Department of Operations Research
Naval Postgraduate School
Code 55JO
Monterey CA 93943
ljohnson@nps-cs.arpa
(408) 646-2569

Johnson Mark A.
MS 7247-267-1
The Upjohn Company
301 Henrietta Street
Kalamazoo MI 49001
(616) 342-9256

Howe Sally
Scientific Computing Division
National Bureau of Standards
Rm A151 Technology Building
Gaithersburg MD 20899
(301) 975-3807

Hsu Jason C.
Department of Statistics
The Ohio State University
141 Cockins Hall
Columbus OH 43210
(614) 292-7663

Hudson Joseph C.
Department of Science and Mathematics
SMI Engineering and Management Institute
1700 W. Third Avenue
Flint MI 48504

Hwang Jimmie
San Diego State Univ.
Social Science Research Lab
San Diego CA 92042
QALJOO2@CALSTATE
(619) 229-2258

Inselberg Alfred
IBM Scientific Center
11601 Wilshire Boulevard
Los Angeles CA 90025-1738
(213) 312-5479

Jernigan Robert W.
Department of Mathematics and Statistics
The American University
4400 Massachusetts Avenue, NW
Washington DC 20016
(202) 885-3120

Johnson Mark
Statistics Group
Los Alamos National Lab
P.O. Box 1663, MS F600
Los Alamos NM 87545
085290CD90@/SIVAX.LANL.GOV
(505) 667-6334

Johnstone Iain
Department of Statistics
Stanford University
Sequoia Hall
Stanford CA 94305
iainj@playfair.stanford.edu
(415) 723-9114

Jost Steve D.
Department of Computer Science
DePaul University
243 S. Wabash Avenue
Chicago IL 60604-2302

Kallianpur Gopinath
Department of Statistics
University of North Carolina
Chapel Hill NC 27514
(919) 962-2187

Kask Alex W.
28-14 210 Street
Bayside NY 11360
(212) 830-5656

Kauffman Tom
Statistics Department
Rice University
P.O. Box 1892
Houston TX 77251
tom@stat5.rice.edu
(713) 975-1173

Keller-McNulty Sallie
Department of Statistics
Kansas State University
Dickens Hall
Manhattan KS 66506
(913) 532-6883

Kettenring Jon R.
BELLCORE
435 South Street, 2Q-326
Morristown NJ 07901
(201) 829-4398

Khoshgoftaar Taghi M.
Department of Computer Science
Florida Atlantic University
P.O. Box 3091
Boca Raton FL 33431-0991
(305) 393-3855

Kipnis Victor
Department of Economics
University of Southern California
MC-0035
Los Angeles CA 90089-0035
(213) 743-2487

Kafadar Karen
Hewlett-Packard
1501 Page Mill Road., 4U
Palo Alto CA 94306
kk%hpspd@phlabs.hp.com
(415) 857-7063

Kanabar Vijay
Department of Mathematics
University of Winnipeg
515 Portage Avenue
Winnipeg Man. R3B 2E9
CANADA
204 786 9345

Katzoff Myron
National Center for Health Statistics
3700 East-West Highway
Hyattsville MD 20782
(301) 436-7047

Keegel John C.
University of DC
1740 Hobart St. NW
Washington DC 20009
(202) 265-2455

Keough Gary
South Building, Rm 4801
USDA/NASS
14th and Independence Avenue, SW
Washington DC 22193
(202) 475-5918

Khan Aqeel A.
Math/Stat Tea, CSCA:RSA
U.S. Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda MD 20814-2797
(301) 295-5566

Kiemele Mark
Air Force Academy
USAFA/DFMS
Colorado Spring CO 80840
(719) 472-4470

Kitagawa Genshiro
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu Minato-Ku
Tokyo 106
JAPAN

Knaub, Jr. James
EI-541, U.S. Department of Energy
Energy Information Administration
1000 Independence Avenue
Washington DC 50585
(202) 586-9619

Korin Basil P.
Mathematics Department
American University
Washington DC 20016
(202) 885-3120

LaVarnway Gerard T.
Department of Mathematics
Norwich University
Northfield VT 05663
(802) 485-5011

Lacayo, Jr. Herbert
U.S.E.P.A. (PM-223)
401 M Street
Washington DC 20480
(202) 382-2714

Landauer Christopher
679 Loring Avenue
Los Angeles CA 90024
(213) 336-5635

Launer Robert L.
614 New Kent Place
Cary NC 27511

Le Hung Tri
Center for Computational Statistics
George Mason University
Fairfax VA 22030
(703) 764-6181

LePage Raoul
Department of Statistics & Probability
Michigan State University
East Lansing MI 48824
RDL@LEPAGE-SUN.STT.MSU.EDU
(517) 353-3984

Lebow William
BBN Software Products
10 Fawcett Street
Cambridge MA 02238
lebow@spca@bbn.com
(617) 873-8128

Lenk Peter
New York University
100 Trinity Place
New York NY 10006
(212) 982-2228

Levene Howard
Columbia University
New York NY 10027
(212) 280-5370

Li Tze Fen
Department of Mathematics
Rutgers University at Camden
Camden NJ 08102
(609) 757-6439

Liggett Walter
National Bureau of Standards
Administration Bldg., A337
Gaithersburg MD 20899
(301) 975-2851

Lii Keh-Shin
Department of Statistics
University of California, Riverside
Riverside CA 92521
(714) 787-3836

Lilliefors Hubert
Statistics Department
George Washington University
Washington DC 20052
HWL1@GWUVM
(202) 994-6664

Lim Dr. Yong Bin
Department of Experimental Statistics
Louisiana State University
Baton Rouge LA 70803
XST720@LSUVM
(504) 388-8383

Lin Charles
2201 Middlefield Court
Raleigh NC 27615
(919) 467-8000

Lin Chih C.
Room 18B-45
FDA, Mail Code HFN-713
5600 Fishers Lane
Rockville MD 20857
(301) 443-4136

Link William
Patuxent Wildlife Research Center
USFWS
Laurel MD 20708

Lippman Alan
Department of Applied Mathematics
Brown University
Box F
Providence RI 02912
(401) 273-4345

Londhe Anil R.
5 Research Parkway
Bristol-Myers Company
P.O. Box 5100
Wallingford CT 06492
(203) 284-6241

Low Leone
Math and Statistics Department
Wright State Unviversity
Dayton OH 45435
(513) 873-2256

Lucke Joseph
Lehigh Valley Hospital Center
1200 S. Cedar Crest Blvd.
Allentown PA 18103
(215) 776-8889

Maar Dr. James R.
National Security Agency
Attn: R51
Fort Meade MD 20755-6000
(301) 859-6341

MacGibbon-Taylor Brenda
Department of Decision Science
Concordia University
1455 de Maisonneuve Blvd. W
Montreal QUE H3G 1M8
Canada
(514) 848-2982

Makris Nick
Gillette Research Institute
1413 Research Boulevard
Rockville MD 20850
(301) 738-0272

Makuch William M.
Research and Development Center
General Electric-CRD
P.O. Box 8 K1-4C41A
Schenectady NY 12301
ARPANET: MAKUCH@GE-CRD
(518) 387-5918

Manchester Lise
Department of Math, Statistics & CS
Dalhousie University
Halifax Nova Scotia B3H 3J5
CANADA
lise@dalcs.uucp
(902) 424-3624

Mann Nancy R.
Department of Biomathematics
University of California
Los Angeles CA 90024

Mathieu Claire
Computer Science Department
Princeton University
Princeton NJ 08544
(609) 452-5496

Matthews Peter
Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore MD 21228
(301) 455-2423

Mazur Catherine
USDA-NASS
Room 4801 S. Bldg.
Washington DC 20250
(202) 475-3483

Mazzuchi Thomas A.
Department of Operations Research
George Washington University
Washington DC 20052
(703) 994-7514

McClave Jim
P.O. Box 14545
Gainesville FL 32605
(904) 375-7624

McClure Don E.
Department of Applied Mathematics
Brown University
Box F
Providence RI 02912
dem@brownvm.bitnet
(401) 863-1496

McDonald Bill
Naval Surface Warfare Center
White Oak Lab
Silver Spring MD 20903-5000
(202) 394-2585

McDonald John Alan
Statistics, GN-22
University of Washington
Seattle WA 98195
JAM6ENTROPY.MS. WASHINGTON.EDU
(206) 545-7438

McIntosh Allen
Room 2C334
Bell Communications Research
435 South Street,
Morristown NJ 07960-1961

McKenzie, Jr. John D.
Babson College
Babson Park
Wellesley MA 02157-0901
(617) 239-4479

McKinney Steve
Department of Biostatistics SC-32
University of Washington
Seattle WA 98195

McLeish D. L.
University of Waterloo
Waterloo Ontario N2L 2V5
CANADA

McLeish Mary
Department of Computing and Information
University of Guelph
Geulph Ontario N1G 2W1
CANADA
(519) 824-4120

Mehra Munish
Department of Statistics
University of Kentucky
Lexington KY 40506
munish@ukma
(606) 257-4423

Mellor-Crummy John
Department of Computer Science
University of Rochester
Rochester NY 14627
crummey@rochester.edu
(716) 275-0922

Mergerson James W.
USDA, NASS, RAD, SSB
3251 Old Lee Highway, Rm. 506
Fairfax VA 22030

Messer Karen
Department of Mathematics
California State University
Fullerton CA
(714) 774-3631

Meyer Michael
Statistics Department
Carnegie-Mellon University
Pittsburgh PA 15213
MM8S@andres.cmv.edu
(412) 268-3108

Meyer Ruth K.
Department of Business Computer Inf. Sys
St. Cloud State University
St. Cloud MN 56301
(612) 255-2241

Mikhail Nabih
Department of Mathematics
Liberty University
Box 20000
Lynchburg VA 24506-8001
(804) 237-5961

Millar P.W.
Department of Statistics
University of California, Berkeley
Berkeley CA 94720
(415) 642-2781

Miller John
Center for Computational Statistics
George Mason University
4400 University Drive
Fairfax VA 22030
(703) 323-2733

Miller Michael F.
Hoechst-Roussel Pharmaceuticals Inc.
Route 202-206N, Bldg M
Somerville NJ 08876
(201) 231-3486

Minor James M.
Du Pont Company
P.O. Box 6090
Newark DE 19714-6090
(302) 366-2432

Mitchell Toby
Math Stat Research, Eng Physics Division
Oak Ridge National Laboratory
P.O. Box Y, Bldg. 9207A, MS3
Oak Ridge TN 37831
mitchell@msr.epm.ornl.gov
(615) 574-3143

Mittnik Stefan
Department of Economics
State University of NY at Stony Brook
Stony Brook NY 11794-4384
SMITTNIK@SBCCVM (Bitnet)
(516) 632-7532

Miyashita Dr. Junryo
Department of Computer Science
California State University
5500 University Parkway
San Bernardino CA 92373
(714) 887-7647

Modarres Reza
Mathematics, Statistics and CS Dept.
The American University
Massachusetts Avenue
Washington DC 20016
(202) 885-3149

Mode Dr. Charles J.
Department of Math. and Computer Sci.
Drexel University
Philadelphia PA 19104
215-895-2668

Mohns Cynthia
Numerical Algorithms Group Inc.
1101 31st Street, Suite 100
Downers Grove IL 60515
(312) 971-2337

Mokatrin Ahmad
Math/Stat Department, Clark Hall
American University
4400 Massachusetts Avenue
Washington DC 20010

Moore Marc
Ecole Polytechnique Montreal
C.P. 6079 Succursale "A"
Montreal Quebec H3C 3A7
CANADA
(514) 340-4513

Moser Barry
Department of Exper. Statistics
Louisiana State University
Ag. Administration Building
Baton Rouge LA 70803-5606
XST769@LSUVM.bitnet
(504) 388-8303

Moussa-Hamouda Effat
DePaul University
2323 N. Seminary, #572-C
Chicago IL 60614
(312) 341-8250

Muller Mervin E.
Department of Computer & Inf. Science
The Ohio State University
2036 Neil Avenue Mall
Columbus OH 43210-1277
(614) 292-5973

Munson Peter J.
National Institutes of Health
Building 10, Room 6C101
Bethesda MD 20892
PTM@NIHCUDEC
(301) 496-2972

Nadas Arthur
IBM Research
P.O. Box 218
Yorktown Height NY 10598
NADAS@YKTVMX
(914) 945-2163

Nelder J.A.
Imperial College
Huxley Bldg., 180 Queens Gate
London SW7 2BZ
ENGLAND
UK+1-589-5111

Newton H. Joseph
Department of Statistics
Texas A&M University
College Station TX 77843-3143
(409) 845-3141

Nguyen Trung H.
EIC Labs
111 Downey Street
Norwood MA 02062
(617) 769-9450

Nicoll Jeff
IDA
1801 N. Beauregarde Street
Alexandria VA 22311
(703) 478-2987

Nolan James R.
Quantitative Business Analysis
Siena College
Loudonville NY 12211
(518) 783-2503

Normolle Daniel
Department of Biostatistics
University of Michigan
School of Public Health
Ann Arbor MI 48109-2029
313-936-1004

Nychka Douglas
Department of Statistics
North Carolina State University
Box 8203
Raleigh NC 27695-8203
NYCHKA@NCSUSTAT.BITNET
(919) 737-2534

O'Brien Fanny
BBN Software Products
10 Fawcett Street
Cambridge MA 02238
(617) 497-3778

O'Connor Carol
Speed Scientific School
University of Louisville
Louisville KY 40292
(502) 588-6304

O'Connor Thomas A.
O'Connor and Associates, Inc.
3017 Juniper Hill Road
Louisville KY 40206

Oldford R. Wayne
Department of Statistics & Act. Sci
University of Waterloo
Waterloo Ontario N2L 3G1
CANADA
CSNET:RWOLDFORD@WATER.WATERLOO.EDU
(519) 888-4609

Ondrasik John A.
P.O. Box 368
Boehringer Ingelheim Pharmaceuticals
90 East Ridge
Ridgefield CT 06877
(203) 798-4243

Ostrouchov George
Mathematical Sciences Section
Oak Ridge National Laboratory
Building 9207A MS-3
Oak Ridge TN 37831
OST@MSR.EPM.ORNL.GOV
(615) 574-3137

Owen Art
Department of Statistics
Stanford University
Sequoia Hall
Stanford CA 94305
art@playfair.stanford.edu
(415) 725-2232

Ozga Martin
Room 4168
USDA, NASS, RAD, SRB
14th & Independence Ave., SW
Washington DC 20250
(202) 447-5483

Pacheco Nelson S.
1259 Lake Plaza Drive
Colorado Spring CO 80906
(719) 550-6376

Padgett William J.
Department of Statistics
University of South Carolina
Columbia SC 29208
(803) 777-5070

Park Tae-woo
AFAL/SCS
Edwards AFB CA 93523-5000
(805) 275-5196

Parzen Emanuel
Department of Statistics
Texas A&M University
College Station TX 77843-3143
(409) 845-3188

Patterson David
Department of Math Sciences
University of Montana
Missoula MT 59812
(406) 243-6748

Peck Roger W.
Department of Computer Science and Stat.
The University of Rhode Island
Tyler Hall
Kingston RI 02881-0816
(401) 792-4497

Pederson Shane
P.O. Box 1663
Los Alamos National Laboratory
Mail Stop F-600
Los Alamos NM 87544
100176A26@S1VAX.LANL.GOV
(505) 667-7303

Pei Gabriel P.
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria VA 22311
(703) 578-2882

Peierls Ronald F.
Applied Mathematics Department - 515
Brookhaven National Laboratory
Upton NY 11973
peierls@bnl.bitnet
(516) 282-4104

Percival Don
Applied Physics Lab
University of Washington
Seattle WA 98105
(206) 543-1300

Perry Charles R.
USDA/NASS
14th & Independence Avenue S.W.
Washington DC 20250
(202) 475-3075

Peruggia Mario
Department of Statistics
Carnegie-Mellon University
232 BH
Pittsburgh PA 15213
(412) 268-8590

Pesek John
College of Agricultural Sciences
University of Deleware
Department of Food & Resource Economics
Newark DE 19717-1303
FOD31626@UDACSVM bitnet
(302) 451-1319

Phillips Abraham
Research Dept
Prudential Property & Cas Insurance Co.
23 Main Street
Holmdel NJ 07733
(201) 946-5109

Phinney Stephen E.
250/060
IBM Manassas
9500 Godwin Drive
Manassas VA 22110
(703) 367-2403

Pickle Dr. Linda W.
NCI
Landon Building, Room 3A06
Bethesda MD 20892
(301) 496-6425

Pieper Carl
295 Central Park West, 15D
New York NY 10024
(212) 799-8212

Pierce Alan
Amoco Production Company
4502 East 41st Street
Tulsa OK 74136
(918) 660-3830

Pierce Margaret Anne
Georgia Southern College
LB 8093
Statesboro GA 30460
(912) 681-5427

Pierchala Carl E.
Food & Drug Administration
P.O. Box 1554
West Bethesda MD 20817
(301) 443-4594

Pitcher Hugh M.
US - EPA
3031 Beech Street, N. W.
Washington DC 20015
(202) 382-2788

Pregibon Daryl
AT&T Bell Labs
600 Mountain Avenue, Room 2C-264
Murray Hill NJ 07974
(201) 582-3193

Puri Madan
Department of Mathematics
Indiana University
Bloomington IN 47405

Puri Madan
Department of Mathematics
Indiana University
Bloomington IN 47405

Raj Baldev
School of Business and Economics
Wilfrid Laurier University
Waterloo Ont N2L 3C5
CANADA
(519) 884-1970

Ramirez Donald E.
Department of Mathematics
University of Virginia
Math-Astro Building
Charlottesville VA 22903
DER@VIRGINIA
(804) 924-4934

Ratnaparkhi M.V.
Department of Mathematics and Statisitcs
Wright State University
Dayton OH 45435
(513) 873-2193

Raubertas Richard F.
Department of Statistics
University of Rochester
Rochester NY 14627
(716) 275-2406

Rayens William S.
Department of Statistics
University of Kentucky
859 Patterson Office Tower
Lexington KY 40506-0027
(606) 257-7061

Richardson David
10211 Brunswick Ave.
Silver Spring MD 20902
(301) 649-2650

Richter Don
Graduate School of Business Admin.
New York University
100 Trinity Place
New York NY 10006
(212) 285-6130

Ringeisen Richard D
Department of Mathematical Sciences
Clemson University
Clemson SC 29634-1907
RDRNG@Clemson
(803) 656-5245

Rom Dror
441 Tomlinson Road #F-13
Philadelphia PA 19116
(215) 661-6336

Rosenblatt Joan R.
National Bureau of Standards
Administration Bldg, Room A438
Gaithersburg MD 20899
975-2733

Rovine Michael J.
Department of Individual and Family Stud
Penn State
S-110 Henderson Building
University Park PA 16802
(814) 863-0267

Rumpf David L.
Department of Ind. Eng. and Inf. Systems
Northeastern University
360 Huntington Avenue, Rm 330 Snell
Boston MA 02115
(617) 437-3632

Ruskin David
Center for Naval Analyses
P.O. Box 16268
Alexandria VA 22302
(703) 824-2284

Russell Carl T.
U.S. Army OTEA, CSTE-TS-R
5600 Columbia Pike
Falls Church VA 22041
(703) 756-1818

Rust John
University of Wisconsin
1180 University Drive
Madison WI 53706
(608) 263-3871

Ryan Barbara
Statistics Department
Stanford University
Sequoia Hall
Stanford CA 94305-4065
BFR@PLAYFAIR.STANFORD.EDU
(415) 723-2787

Ryan Thomas A.
Minitab, Inc.
3081 Enterprise Drive
State College PA 16801
(814) 238-3280

Salomon Matthew A.
Fiscal Analysis Division
Congressional Budget Office
2nd & D Streets, SW
Washington DC 20515
(202) 226-2765

Samuels Stephen M.
Department of Statistics
Purdue University
W. Lafayette IN 47907
(317) 494-6042

Samuelson Douglas A.
International Telesystems Corporation
600 Herndon Parkway
Herndon VA 22070

Sawyer John W.
Texas Tech University
1A104 Health Sciences Center
Lubbock TX 79430
(806) 743-2146

Scharff H. Felix
Dept 49WA, Stu 260
IBM
Neighborhood Road
Kingston NY 12401
(914) 385-4013

Schiller Susannah
5 Scharlet Sage Court
Burtonsville MD 20866
Schill@NBS
(301) 776-1187

Schmee Josef
Union College
Bailey Hall 311
Schenectady NY 12308
schmeej@union
(518) 370-6248

Schmeiser Bruce W.
School of Industrial Engineering
Purdue University
Grissom Hall
West Lafayette IN 47907
schmeise@pink.ecn.purdue.edu
(317) 494-5422

Schwemberger John
8154 Larkin Lane
Vienna VA 22180
(202) 382-7195

Scott David
Department of Statistics
La Trobe University
Bundoora Victoria 3083
AUSTRALIA
STADTS@latvaa8.lat @murnari.oz
+(61)3 4792091

Scott David W.
Department of Statistics
Rice University
P.O. Box 1892
Houston TX 77251-1892
(713) 527-8101

Scott Dr. David
Dept of Decision Sciences and Management
Concordia University
1455 de Maisonneuve Blvd. West
Montreal Quebec H3G 1M8
CANADA
(514) 848-2969

Seaman John W.
Department of Statistics
University of Southwestern Louisiana
P.O. Box 41006
Lafayette LA 70506-1006
(318) 231-5294

Segall Dr. Richard
Department of Mathematics
University of Lowell
Lowell MA 01854
(617) 452-5000

Sessions Dr. David N.
9 Purdue Road
Glen Cove NY 11542
(516) 676-2123

Shanmugam Ron
University of Colorado
Denver CO 80204
(303) 556-8463

Shiau Jyh-Jen Horng
Department of Statistics
University of Missouri-Columbia
222 Mathematical Science Bldg
Columbia MO 65211
(314) 882-7467

Shiau Tzong-Huei
Department of Computer Science
University of Missouri-Columbia
Columbia MO 65211
csshiau@umcvmb.bitnet
(314) 882-4540

Shier Douglas R.
Department of Mathematics
College of William and Mary
Williamsburg VA 23185
SDRSHIE@WMMVS
(804) 253-4481

Shing Chen-Chi
Department of Computer Science
Radford University
Box 5752
Radford VA 24142
(703) 831-5733

Shrager Richard I.
National Institutes of Health
Bldg 12A, Room 2041
Bethesda MD 20892
(301) 496-1122

Silverberg Arthur
Food and Drug Administration
4600 Fishers Lane, HFV-124
Rockville MD 20857
(301) 443-1580

Simon Dr. Richard
Room 4B06
National Cancer Institute
7910 Woodmont Avenue
Bethesda MD 20892
(301) 496-4836

Singpurwalla N.D.
Department of Operations Research
George Washington University
Washington DC 20052
(202) 994-7515

Siu Cynthia O.
Osler 622
The Johns Hopkins University
600 N. Wolfe Street
Baltimore MD 21205

Sleeper Lynn A.
Department of Biostatistics
Harvard School of Public Health
677 Huntington Avenue
Boston MA 02115
(617) 732-3626

Slinkman Craig W.
Department of Inf Sys & Mgmt Sciences
The University of Texas at Arlington
Arlington TX 76019
B718MEE@UPARLVM1
(817) 273-3502

Slowinski Samuel M.
Federal Reserve Board
20th & Constitution Avenue, NW
Washington DC 20551
(202) 452 2622

Smith Adrian
Department of Mathematics
University of Nottingham
Nottingham NG7 2RD
UNITED KINGDOM
602 484848

Smith Eric P.
Department of Statistics
Virginia Polytechnic Institute & State U
Blacksburg VA 24061
(703) 961-7932

Smith Laurie Melany
9521 Baltimore Avenue
Laurel MD 20707
(301) 490-9665

Snell Robert
Eastman Kodak Company
MQA, Bldg 56, Fl 4, KP
Rochester NY 14650

Sofer Ariela
ORAS
George Mason University
4400 University Drive
Fairfax VA 22030
(703) 323-2728

Somerville Paul N.
University of Central Florida
P.O. Box 25000
Orlando FL 32816
(305) 275-2695

Song Wheyming Tina
Purdue University
West Lafayette IN 47906
(317) 698-7961

Sood Arun K.
Department of Computer Science
George Mason University
4400 University Drive
Fairfax VA 22030
(703) 323-3395

Soyer Refik
Department of OR
George Washington University
Washington DC 20052
(202) 994-6794

Speckman Paul
Department of Statistics
University of Missouri-Columbia
Columbia MO 65211
(314) 882-7783

Stephenson Elizabeth
P.O. Box 7375
2375 Garcia Avenue
Moutain View CA 94039
(415) 960-7784

Stern Hal
Department of Statistics
Harvard University
One Oxford Street
Cambridge MA 02138

Stewart G. W.
University of Maryland
College Park MD 20742
stewart@thales.cs.umd.edu
(301) 454-6120

Stewart Leland
Department 92-20, Bldg. 254E
Lockheed Research Laboratory
3251 Hanover Street
Palo Alto CA 94304
(415) 424-2710

Studdiford Walter B.
Registrar's Office
Princeton University
B10 A West College
Princeton NJ 08544
(609) 452-6195

Stuetzle Werner
Department of Statistics
University of Washington
GN-22
Seattle WA 98195
(206) 543-4386

Sutton Cliff
Center for Computational Statistics
George Mason University
242 Science Technology Building
Fairfax VA 22030
csutton@gmuvax
(703) 323-3863

Szewczyk William F.
2905 Shamrock Terrace
Olney MD 20832
(301) 774-1158

Takane Yoshio
Department of Psychology
McGill University
1205 Dr. Penfield Avenue
Montreal Quebec H3A 1B1
CANADA
PS81@MCGILLA
(514) 398-6125

Tarter Michael E.
Department of Statistics
University of California
32 Earl Warren Hall
Berkeley CA 94720
(415) 642-4601

Tasker Gary D.
U.S. Geological Survey
430 National Center
Reston VA 22092
(703) 648-5892

Tawfik Lorraine
40 Amityville Street
Islip Terrace NY 11752
(516) 277-2875

Taylor D. Wayne
Department of Clinical Epidemiology&Bio.
McMaster University
Health Sciences Center
Hamilton Ontario L8N 3Z5
CANADA
525-9140 X4102

Teitel Robert F.
Teitel Data Systems
7200 Wisconsin Avenue, Suite 410
Bethesda MD 20814
(301) 656-0401

Terpenning Irma
Rd2 Box 109
Frenchtown NJ 08825
(201) 582-2268

Therneau Terry M.
May Clinic
200 First Street SW
Rochester MM 55905
(507) 284-8803

Thisted Ronald A.
Department of Statistics
University of Chicago
5734 University Avenue
Chicago IL 60637
thisted@galton.uchicago.edu
(312) 702-8333

Thompson James R.
Department of Statistics
Rice University
P.O. Box 1892
Houston TX 77251-1892
(713) 527-4828

Thornton Ding H.
Naval Air Test Center
Computer Sciences Directorate
Patuxent River MD 20670-5304
(301) 863-3396

Tierney Luke
School of Statistics
University of Minnesota
Minneapolis MN 55455
luke%umnstat@umn-cs.arpa
(612) 625-7843

Tretter Dr. Marietta
Business Analysis & Research
Texas A.& M. University
College Station TX 77843
(409) 845-1383

Tsong Yi
FDA HFN-715
6500 Fisher Lane
Rockville MD
(301) 443-4710

Tsou Tai-Houn
3637 Canyon Crest Drive, A307
Riverside CA 92507
(714) 788-4656

Tubb Gary
Instructional Computing
University of South Florida
USF 3185
Tampa FL 33620
DNPABAA@CFRVM

Tukey Paul A.
Bell Communications Research
435 South Street
Morristown NJ 07960
(201) 829-4285

Turner David L.
Department of Mathematics & Statistics
Utah State University
Logan Utah 84322-3900
DTURNER@USU
(801) 750-2814

Unger Elizabeth
Computing and Information Sciences
Kansas State University
243 Nichols Hall
Manhattan KS 66506

Utts Jessica
SRI
333 Ravenswood Avenue
Menlo Park CA 94025
utts@unix.sri.com
(415) 859-4445

Varner Ruth
369 Holmes Drive
Vienna VA 22180
(703) 938-9209

Varty John Franklin
6602 Boulevard View Place
Alexandria VA 22307
(703) 765-0540

Venetoulias Achilles
E40-133
MIT, Sloan School
1 Amherst Street
Cambridge MA 02139
axilleas@dolphin.mit.edu
(617) 253-8416

Vernhes Frederique L.
Department of Statistics
Yale University
Box 2179, Yale Station
New Haven CT 06511
(203) 782-0430

Vetter John E.
Washington Navy Yard
Naval Weapons Engineering Support Act
ESA-31, Bldg. 220-2
Washington DC 20374-2203
(202)433-3621

Von Eye Alexander
Department for Individual Family Study
The Pennsylvania State University
University Park PA 16802
(814) 863-0267

Wahba Grace
Department of Statistics
Yale University
Box E2179 Yale Station
New Haven CT 06520-2179
wahba@celray.cs.yale.edu
(203) 432-0666

Wang Chaiho
1232 Meyer Court
McLean VA 22101
(202) 724-6368

Weber James S.
Department of Management
Roosevelt University
P.O. Box 603
Gurnee IL 60031-0603

Weidman Scott
MRJ, Inc.
10455 White Granite Drive
Oakton VA 22124
(703) 385-0879

Weiss Robert E.
Department of Applied Statistics
University of Minnesota
Classroom Office Bldg. 352
St. Paul MN 55108
weiss@umnstat.stat.umn.edu
(612) 625-2756

Wesley Robert
Department of Health and Human Services
9807 Owen Brown Road
Columbia MD 21045
(301) 496-7946

Vitter Jeffrey S.
Department of Computer Science
Brown University
Box 1910
Providence RI 02904
jsv@cs.brown.edu
(401) 863-3300

Waclawiw Myron
5364 Hesperus Drive
Columbia MD 20144
(301) 730-0294

Walker Homer
Department of Mathematics
Utah State University
Logan UT 84322-3900
uf7099@usu.bitnet
(801) 750-2026

Wang R. H.
P.O. Box 586
OLIN
350 Knotter Drive
Cheshire CT 06410
(203) 271-4196

Wegman Edward J.
Center for Computational Statistics
George Mason University
242 Science Technology Building
Fairfax VA 22030
ewegman@gmuvax.gmu.edu
703 323 2723

Weiss Guenter
University of Winnipeg
515 Portage Avenue
Winnipeg Man R3B 2E9
CANADA
(204) 786-9399

Welsch Roy E.
M.I.T.
50 Memorial Drive, E53-383
Cambridge MA 02139
(617) 253-6601

Whitney David A.
TASC
55 Walkers Brook Drive
Reading MA 01867
(617) 942-2000

Whitridge Patricia
Business Survey Methods Division
Statistics Canada
RH Coats Bldg 11-C, Tunney's Pasture
Ottawa Ontario K1A 0T6
CANADA
(613) 951-8614

Wilburn Arthur J.
4600 Jasmine Drive
Rockville MD 20853-1737
(301) 929-1040

Wilson P. David
504 Shadow Grove Court
Lutz FL 33549
(813) 974-4860

Winkler Gernot
Time Service Department
U.S. Naval Observatory
34th & Massachusetts Avenue, NW
Washington DC 20392-5100
(202) 653-1520

Winkler William E.
Census Bureau
Washington DC 20233
(301) 763-3905

Wochnik Michael
1212 Gibbon #6
Laramie WY 82070
(307) 745-9393

Wollan Peter
Department of Mathematics
Michigan Technological University
Houghton MI 49931
USA
(906) 487-2694

Wolting Duane
Aerojet TechSystems Company
P.O. Box 13222, Bldg. 2002, Dept. 9470
Sacramento CA 95813
(916) 355-2692

Woodburn Rose Louise
8426 Ravenswood Road
New Carrollton MD 20784
(301) 459-5138

Woodfield Terry J.
SAS Institute Inc.
SAS Circle, Box 8000
Cary NC 27512-8000
(919) 467-8000

Woodruff Brian
Bolling AFB
AFOSR
Washington DC 20332
(202) 767-5027

Wyscarver Roy A.
Economic Modeling & Computer Application
U.S. Treasury Department
15th & Pennsylvania Ave., NW
Washington DC 20220
(202) 566-5085

Yang C. C.
NRL
Code 5380
Washington DC 20375

Yang Ting
University of Cincinnati
ML 025
Cincinnati OH 45221
(513) 475-5619

Young Dean M.
Department of Information Systems
Baylor University
Waco TX 76798
(817) 755-2258

Youngren Mark A.
3809 Terrace Drive
Annandale VA 22003
(202) 295-1625

Yu Chen Cheng W.
Dept. 23W, Bldg. 630, E60
IBM Corporation - East Fishkill
Route 52
Hopewell Junc. NY 12533-0999
(914) 892-2200

Appendix C
Detailed Expenditures

Interface Conference Expenses Billed to AFOSR

Clerical Support			
Salary to Registration Personnel			1122
Travel			
Hal Stern	116		
Total Travel		116	
Per Diem			
Munish Mehra	334		
M. Bolorforoush	25		
John Miller	31		
Kim Anh Do	334		
Claire Mathieu	69		
Hal Stern	25		
Total Per Diem		818	
Registration Remission			
Jerome Liang	130		
Ahmad Mokatrin	105		
Reza Modarres	105		
Kim Anh Do	105		
Y. B. Lim	105		
Daniel Normolle	105		
Andrew Bruce	105		
Lynn A. Sleeper	95		
Jeff Banfield	105		
Tina Song	105		
Celesta Ball	130		
M. Bolorforoush	105		
Hung Le	105		
John Miller	105		
Tom Kaufman	105		
Douglas Nychka	105		
Claire Mathieu	130		
Bradley Efron	105		
Kathryn Chaloner	130		
R. W. Oldford	105		
Katherine Hurley	95		
Deborah Donnell	120		
Naomi Altman	105		
Hal Stern	95		
Total Registration Remission		2605	
Invited Speaker Honorarium			
Thomas Banchoff	500		
Wolfgang Haerdle	575		
Total Invited Speaker Honorarium		1075	
Total Participant Expenses			4614

Miscellaneous Expenses		
Letterhead	282	
Signs and Signholders	289	
Proceedings Expenses	976	
Certificates	52	
Audio-Visual Rental	1686	
Duplicating	70	
Total Miscellaneous Expenses		3355
Total Direct		9091
Indirect at 10% of Total Direct		909
Grand Total		10000